Robust Dynamic Mechanisms

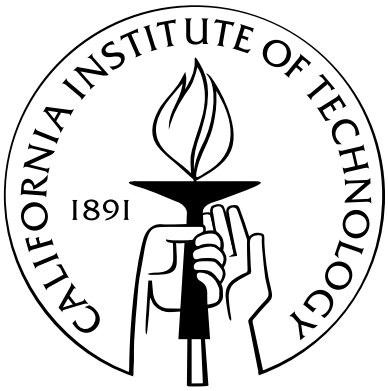
Thesis by

Mohamed Mostagir

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2012

(Defended September 2, 2011)

Acknowledgements

The following people and places have made my time at Caltech an educational and enjoyable experience.

Mentors: John Ledyard, Preston McAfee, Tom Palfrey, and Jean-Laurent Rosenthal. Further advice from Marina Agranov, Colin Camerer, Federico Echenique, Erik Snowberg, Leeat Yariv, and Adam Wierman.

Friends: Greg and Pia, Dustin and Katya, Sera, Ines, Salvo, Julian, Gui, Andrea, Maggie, Boosey, Marjan, Andrej, Ian, Marwa Mabrouk, Dani and Gabri, Lisa and Costia, Xiao Jun.

Friends-at-a-distance: Theo, Katy, Peach, Khaldoun, and Kostas.

Student Life: Tim! Sue, Barbara, John, Geoff, Mike, Juan, and all the wonderful people at Caltech Housing.

Administrative: Edith, Laurel, Suzanne, Gloria, Gail, Sheryl, and Victoria.

The Mediterranean Cafe!

Lloyd House: the highlight of my time at Caltech. I live and die for those I love.

Family: Mesbah, Mona, and Mai.

I have spent the best years of my life in Pasadena. This in no small part was due to the fact that my days were consistently brightened up by Hadil's exuberant presence. Laila Hikaru made her appearance towards the end and life became even more beautiful. I would dedicate this thesis to the two of them, but I will instead save that for my best work, which is still ahead of me.

"Sometimes a scream is better than a thesis" — Ralph Waldo Emerson.

Abstract

This thesis presents and solves two dynamic problems. The first problem comes from online display advertising. In display advertising, a publisher displays an ad for an advertiser when a targeted user visits a webpage related to the advertiser's products or services. However, the publisher cannot control the supply of display opportunities, and hence the actual supply of ads that it can sell is stochastic. I consider the problem of optimal ad delivery, where the advertiser demands a certain number of impressions to be displayed over a certain time horizon. Time is divided into periods, and in the beginning of each period the publisher chooses a fraction of the *still unrealized* supply to allocate towards fulfilling the publisher's demand. The goal is to be able to fulfill the demand at the end of the horizon with minimal costs incurred from penalties associated with shortage or overdelivery of impressions. For a special case of this problem I describe an optimal policy that is very easy to implement. The general version of the problem is more computationally demanding, but I describe policies that are both implementable and arbitrarily close to the optimal solution.

In the second part of the thesis, 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 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. The last part of the thesis considers two more learning

models, best response and fictitious play, and derives the principal's optimal policies theoretically and computationally for the same class of games considered in the social learning model.

Contents

List of Figures

Chapter 1

Introduction

This thesis studies optimization problems in dynamically changing environments. Unlike static optimization problems, where all the relevant information for solving the problem is available to the decision maker in advance, dynamic problems present a myriad of difficulties that arise for a multitude of reasons. Uncertainty about the state of the world presents one such difficulty: the state that the environment is in may be revealed in stages instead of all at once. This requires the decision maker to continuously change their actions in order to respond to a variety of possible scenarios. The complexity of dealing with and responding optimally to such contingencies can be very high, often making it infeasible for the decision maker to develop fully-contingent plans.

Another source of difficulty arises from interacting repeatedly with an opponent. Such interactions require the players to think ahead about the future in order to decide on their best course of action, again taking into account the possibly huge joint action space when charting out their plan. In contrast to static problems, solving a dynamic problem requires the decision maker to not just optimize for today, but to think about how current decisions affect future payoffs. A decision that is optimal for today's problem may not be ideal when one takes the future into account. This tension between short and long-term objectives usually adds to the difficulty of problems that take place in a dynamic setting.

A more optimistic view of dynamic problems is that they provide the decision maker (or the players, in case of a game) with an opportunity to learn about their environment and consequently improve the decisions they make. This is particularly important, not to mention realistic, in environ-

ments that agents find themselves in with no prior experience. The standard approach to decision making in economics assumes that rational agents will always make optimal decisions regardless of the situation they find themselves in or whether they have played the game before. In reality, many games are played repeatedly and it would be expected that agents would behave differently as they become more familiar with the game and with their opponents. In that sense, thinking about dynamic problems becomes not a mere technical curiosity, but a more accurate depiction of various situations that arise daily when agents interact amongst themselves or with their environment.

The thesis examines the preceding issues in the context of two dynamic problems. The first problem comes from the field of online advertising. In this problem, the decision maker is optimizing against an uncertain environment and has to deal with the aforementioned array of difficulties that comes along in such environments. The second problem involves a decision maker, or a principal, who is interacting with a crowd of learning agents in the context of a repeated game. In both problems, my main concern is deriving and understanding the structure of the optimal policies that the principal should use to maximize his payoff. Various related questions are answered as an extension of the main results that address finding the optimal policies: How do these policies compare to other policies that may be less computationally burdensome but only approximate the optimal solution? How does thinking about agents as learning, evolving entities instead of fully rational computing machines change how the principal should play the game? Are there situations that are better described by these models than the standard economics model?

The following is a description of the problems addressed in this thesis, as well as a summary of the results and contribution.

1.1 Display Advertising

In the first chapter I consider an optimization problem that comes up in the field of online display advertising (Ghosh, McAfee, Papineni, and Vassilvitskii 2009). In online display advertising, a publisher targets a specific audience by displaying ads on content web pages. A display opportunity occurs when a member of the target audience visits a webpage that the publisher can post an ad

on. Because the publisher has little control over internet traffic, the supply of display opportunities is stochastic. I consider the problem of optimal ad delivery, where an advertiser requests a number of ads to be displayed by the publisher over a certain time horizon. Time is discrete and divided into periods. In the beginning of each period the publisher chooses fractions of the *still unrealized* supply to allocate towards fulfilling the advertisers' demands. If the publisher fails to deliver the agreed-upon demand at the end of the horizon, it is charged a penalty per each undelivered ad. At the same time, if the publisher supplies more ads than required then there is also a penalty associated with overdelivery. Possible reasons for the existence of such a penalty are given in the next chapter. The goal is to be able to fulfill the demand at the end of the horizon with minimum costs incurred from penalties associated with shortage or overdelivery of ads as well as advertiser-specific delivery constraints.

This is an example of a dynamic problem where the main source of difficulty comes from uncertainty (of the supply). If supply in each period was certain, then the problem would be trivial: the publisher just assigns fractions of the supply in each period until demand is fulfilled, with no risk of running over at any point. The problem becomes trickier when supply is uncertain as there are too many contingencies to plan for and the computational burden becomes too high.

1.1.1 Contribution

There are two main contributions in this chapter:

- The first is isolating a special case of the display advertising problem and fully characterizing the optimal policy, in terms of both its structure and how it can be computed. The optimal policy in this case has a surprisingly simple structure — characterized by a vector of positive numbers, one for each period— that can be efficiently computed and used on the fly at any point in the problem to determine the optimal fraction of supply to assign, regardless of the path that the problem has taken prior to that point.

- When the general case is considered, the problem becomes more difficult to solve optimally, as an algorithm that deals with all possible scenarios will require time that is pseudopolynomial

in the input size (Garey and Johnson (1979)). This means that the complexity of solving the problem directly depends on some of the values in the input (so that for example, a problem where the demand is 1000 is considerably more difficult to solve than one where the demand is 100). To get around this problem, I design a complexity/cost trade-off scheme that allows the publisher to get as close as it wants to the optimal solution at the expense of a more complex problem to solve. Thus a publisher who is looking for a quick solution and doesn't mind an extra bit of expense can approximate the problem more roughly (and incur more cost) than a publisher who is willing to wait for the solution of a more complex problem.

1.1.2 Methodology

The techniques used in this chapter rely on finite horizon dynamic programming. One can show that the cost-to-go/value function is always convex in the state and decision variables. For the special case of the problem, this reduces to finding a solution to a series of disjoint convex minimization problems, which can be done efficiently and allows for nice closed-form solutions for the optimal policy.

While convexity is maintained in the general case, there are no closed-form descriptions of the value function. Instead, the value function has to be constructed for all possible states in each time period. Since the number of states in each period is directly tied to the input of the problem, this makes the problem quite difficult, as the publisher may have to deal with a very large number of states. The approach I take relies on geometric rounding: The state space is divided, or partitioned into regions. Each region is represented by one element from that region. The challenge is doing this representation without breaking down the convexity of the (now approximate) value function so that the minimization problems associated with each period can still be solved efficiently. As the publisher requires more accurate solutions, the state space is divided more finely, allowing for more detail in computing the approximate value function, but at the expense of a larger state space.

1.2 Exploiting Myopic Learning

The second chapter shifts focus from the straightforward dynamic optimization under uncertainty problem to a more game-theoretic setting. In this chapter, I consider a repeated interaction between a principal and a population of learning agents. The learning model considered is that of the replicator dynamics, where agents copy the strategies of their more successful counterparts. I analyze a game, called the Cheat-Audit game, which is a variation on asymmetric matching pennies. The game is played by the principal on one side and the population on the other, and the goal is for the principal to manipulate the learning dynamics to control or limit the fraction of agents taking an action that the principal considers harmful. As I discuss in the chapter, there are a variety of applications of the model, most notable is the one on illegal (music) file sharing.

The driving question behind the work in this part of the thesis is whether it is possible to obtain results that improve on the standard model of decision making in economics when some kind of learning is incorporated into the agents' behavior, so that agents do not immediately respond to changes in the environment, but instead there is some lag between when a certain action is taken and the time most of the population starts responding optimally to this action.

This dynamic aspects of this problem are different from those in the first chapter. There is no uncertainty here as I consider an infinite population of agents and state transitions take place with probability one. The principal's actions reverberates through the population through social learning, which takes some time to happen, and the main difficulty comes from the tension between optimizing for the current period and optimizing for the future.

1.2.1 Contribution

There are three main contributions in this part of the thesis

- I show that by understanding the dynamics of the population and taking the future into account, the principal can sometimes obtain a higher payoff than that of the Nash equilibrium while exerting less effort than what the Nash solution requires. This is a constructive result,

meaning that I do not just show that it is possible for the principal to obtain better payoffs, but I give a detailed description of how he should play the game to guarantee such payoffs.

- I provide practical examples that show that the standard way in which the game is played, with everyone being fully rational, is not always a good description of reality, and that the learning model I use is able to provide a better explanation of such examples.

- On the conceptual front, I argue that imperfect decision making in a population —as exemplified by learning— can in some cases be considered a resource that most system planners fail to utilize.

1.2.2 Methodology

This part of the thesis uses methods from optimal control theory. I derive the optimal policy for the principal via the use of Hamiltonian and variational calculus techniques. These techniques provide necessary but not sufficient conditions that an optimal policy should fulfill. I prove the existence of an optimal policy and use the necessary conditions to show that the policy derived is unique.

When considering the case for a myopic population *and* a myopic principal, the equations of motion that describe the evolution of each party's actions constitute a dynamical system that has a unique non-hyperbolic equilibrium. I solve the dynamical system by examining the Hartman-Grobman linearization of the Jacobian of that system near the equilibrium.

1.3 Best Response and Fictitious Play

The last part of the thesis extends the work on learning to Best Response and Fictitious Play models, while still considering the same Cheat-Audit game. In contrast with the chapter on replicator dynamics, I consider a discrete time setting. When agents use best response, they act as if the principal's action in the most recent period of play is the same action that he will take in the current period, and they best respond to that action. Under fictitious play, the agents respond to the average of the principal's play over time. This implies the possibility of manipulation, since the principal

can conceal temporary deviations from his course of play in the overall average of past play, leading to potential gains in payoff.

The question is indeed whether one can obtain similar results to the ones obtained under the replicator dynamics model. The answer turns out to be mixed: while it is indeed possible in most situations to improve on the Nash solution, the resulting policies are very sensitive to parameter values, such that very slight differences in value can lead to a complete change in policy.

1.3.1 Contribution

The contributions in this chapter are both theoretical and computational:

- On the theory side, I characterize the optimal policy for the principal when agents are using best response. I show that depending on the parameters of the problem, the policy either alternates between periods of auditing and not auditing, or audits at a constant low rate. In either case, the principal can always do better than the Nash solution.

- I computationally find the optimal policy when agents are learning according to fictitious play, and show that there is a strong resemblance to the best response optimal policies. In particular, one candidate for the optimal policy is a threshold strategy, where the principal only audits the entire population when the history of auditing becomes weak (as in, according to history, there has not been too much recent audit activity). Whether the principal can perform better than the Nash solution depends on the parameters of the problem. In particular, the cost in the Nash solution does not depend on the Nash audit rate whereas in fictitious play it plays a major role in determining the optimal policy and its cost.

1.3.2 Methodology

The theoretical methods used in this chapter come from infinite horizon dynamic programming, where a candidate policy is examined for optimality by checking for profitable deviations or lack thereof. The conditions for which a policy is optimal come from restrictions placed on some of the parameters so that no profitable deviations exist.

The computational part solves a finite version of the dynamic program and approaches the infinite horizon case for a relatively low number of periods. The dynamic program is solved via MATLAB and the code is provided in the appendix.

Chapter 2

Optimal Delivery in Display Advertising

2.1 Introduction

Display advertising has become one of the most profitable areas of online services, responsible for approximately $24 billion in business (Ghosh, McAfee, Papineni, and Vassilvitskii 2009). Unlike sponsored search, where textual ads are displayed along the results of a keyword search, display advertising targets specific audiences by showing graphical banner ads on regular content pages. Targeting can be specific by focusing on certain demographics, so that for example, an ad is only shown to people from a certain age group living in a particular geographic location. Typically, display advertising is handled through direct contracts between the publisher and the advertiser. These contracts are characterized by the publisher committing to the delivery of a pre-specified number of ads to the target audience during a certain time period. Because the supply of display opportunities is uncertain, it is possible that the publisher is unable to fully meet the advertiser's demand, in which case the advertiser is compensated via a penalty (per undelivered impression, for example). Additionally, overdelivering, or providing an advertiser with more impressions than their requested demand can be costly for a variety of reasons.[1] The tension between the shortage and overdelivery costs in addition to the stochasticity of the supply is what makes the publisher's

[1]For example, there may be an opportunity cost associated with giving the ad away instead of selling it to another advertiser. It is also possible that the advertiser's infrastructure can only handle so many visits from people who see the ad and click on it before that infrastructure breaks down, and so a cap is placed on the number of ads that the advertiser wants displayed during a period of time.

problem difficult. The basic question I deal with in this chapter is the following: Given an advertiser's demand, a finite planning horizon, and a time-variable supply distribution, how should the publisher dynamically choose fractions[2] of the *still unrealized* supply in each period so that the total expected cost is minimized under the various penalties?

As in other forms of online advertising, ads are assigned to advertisers through the use of auctions. Because of the intricacies and complexities of these auctions and the overhead required by the advertisers to handle them, many advertisers simply opt to let the publisher manage their campaigns and do their bidding on their behalf. As in Feige, Immorlica, Mirrokni, and Nazerzadeh (2008), the advertiser indicates a maximum price that it is willing to pay per impression, and the publisher uses this constraint when bidding on impressions for the advertiser. With the volume of traffic generated over the internet, these auctions take place at an extremely fast rate. It would thus be inefficient, if not completely impossible, to adjust the advertiser's bid after every single auction. Therefore, the advertiser's bid, placed by the publisher, remains effective for a certain period of time until it is re-adjusted for the next time period. By having a constant bid placed over all the auctions taking place in a time period, one can expect to win a fraction of these auctions. I will make use of this correspondence between bids and fractions in my formulation by thinking of the decision variables as fractions of the uncertain supply instead of bid values for each time period. This has been the standard approach in recent work on the problem (e.g., Boutilier, Parkes, Sandholm, and Walsh (2008) and Ghosh, McAfee, Papineni, and Vassilvitskii (2009)). Like these papers, I think of the supply of ads as a 'channel' with an uncertain capacity. However, unlike the area of literature that focuses on selecting the optimal set of contracts to maximize revenue in such a setting (for example, Babaioff, Hartline, and Kleinberg (2008), Constantin, Feldman, Muthukrishnan, and Pál (2008), and Feige, Immorlica, Mirrokni, and Nazerzadeh (2008)), I take the contract as input and focus on how to *optimally* fulfill the demand under supply uncertainty. I assume that the only control that a publisher exerts over the supply is to decide on a fraction of the channel to allocate towards fulfilling an advertiser's demand before the actual supply is realized for that period. Instead of formulating

[2]The reason our decision variables are fractions of the supply will be clear shortly.

the problem as that of profit maximization —by fulfilling as much as possible of the demand for the negotiated price per impression— I think of it as a cost minimization problem, where one tries to minimize the number of ads not served (equivalent to lost revenue in the maximization model) in addition to the overdelivery penalty discussed earlier. The main question I am interested in here is similar to some of the questions asked in Boutilier, Parkes, Sandholm, and Walsh (2008). There, the authors aimed to give a very general, all-encompassing framework to the problem at the expense of giving solutions that provide no performance guarantees. In contrast to their work, I focus on the specific problem described above and I am able to completely characterize the optimal policy under reasonable assumptions. I also show that while we cannot obtain such a solution for the general case, we can get arbitrarily close to the optimal solution.

Our understanding of online advertising has evolved from looking at the problem as a sequence of seemingly unrelated single-round auctions to become more of a carefully planned campaign that admits more expressive requests from the advertiser's side. For example, as noted earlier, advertisers can be very specific in defining their target groups. In addition, there can be other side constraints or terms added to the publisher's contract. As an example, a contract can specify that, in addition to requiring a certain number of impressions to be delivered over a period of thirty days, the delivery should also be spread as evenly as possible, so that if the demand is, say, 300,000 impressions, then the advertiser would ideally prefer to display 10,000 impressions every day for the duration of the contract. This way the advertiser gets a more steady exposure instead of a possible burst in delivery followed by no advertising that the earlier setting allows (for example, by delivering all ads on the first day and then doing nothing for the rest of the planning horizon). One can easily imagine many ways in which the advertiser can amend their contract to include constraints like the above example. I will give a sufficient condition under which the methods in this chapter extend to more expressive contracts.

There is a strong connection between the problems in this chapter and problems from the theory of stochastic inventory control. The literature in this area is vast, with a standard model of stochastic demand (see Zipkin (2000)) but scattered and problem-specific models for random yield

and stochastic supply (Yano and Lee (1995)). Until recently, the focus of this literature has been on identifying the structure of the optimal policies for these problems without much regard to the feasibility of actually computing such policies. Most of these policies were based on dynamic programming formulations and solving the dynamic program was costly and in many cases impossible. Later work was successful in finding approximate policies that either do not rely on dynamic programming, for example, Levi, Pal, Roundy, and Shmoys (2007) or that exploit the structure of the dynamic program to provide near-optimal solutions without the computational burden (Halman, Klabjan, Mostagir, Orlin, and Simchi-Levi (2009)).

The rest of the chapter is organized as follows. Section 2.2 gives a formal definition of the problem, while Section 2.3 derives the optimal policy for a special but important case. Section 2.4 derives an approximation scheme for the general case. Section 2.5 shows how to extend the solution to the case with multiple advertisers as well as extensions to more expressive contracts. Section 2.6 concludes the chapter and suggests possible extensions to the results obtained herein.

2.2 Model and Notation

I will highlight some of the methods used throughout the chapter by focusing on the single-advertiser case for most of this section, and so I present the model for this case first. The extension to multiple advertisers is given in Section 2.5. First, consider the demand side of the problem. An advertiser requests a number of ads that it would like displayed over a certain time horizon. Time is discrete and is divided into periods, with the planning horizon consisting of T periods. The advertiser wishes to have a total of D impressions delivered over the entire horizon. Later on I will discuss the case when the advertiser can also specify additional requirements, like even spacing of impressions over time, etc.

The supply is stochastic and time-variable. In each time period $t, t = 1, ..., T$, the publisher gets a random number X_t of display opportunities that are related to the advertiser's target group. Here, X_t is a random variable that is distributed according to a known distribution $F_t(x)$, with density $f_t(x)$. We assume that the supply distributions across periods are independent, but not necessarily

identically distributed. In each period t and before X_t is realized, the publisher decides on a fraction $\alpha_t, 0 \leq \alpha_t \leq 1$, to be taken out from the random supply X_t in order to fulfill part of the demand D. As discussed earlier, this fraction is equivalent to selecting a bid that ultimately awards the advertiser a fraction of the supply at the end of the period. At the end of the planning horizon, the publisher incurs a penalty per undelivered impression, denoted by p_1. There is also a penalty per overdelivered impression, which can be thought of as the cost of giving away an impression for free instead of selling it. We will denote this penalty by p_2. At time $T = 0$, the expected cost over the planning horizon can be expressed by the following loss function

$$\mathcal{L}(D, \alpha) = E\left[p_1\left(D - \sum_{t=1}^{T} \alpha_t X_t\right)^+ + p_2\left(\sum_{t=1}^{T} \alpha_t X_t - D\right)^+\right],$$

where $(y-a)^+ = \max(y-a, 0)$. The publisher's problem is to select the fractions $\alpha_1, ..., \alpha_T$ such that $\mathcal{L}(D, \alpha)$ is minimized. Put differently, the publisher wants to find a policy whereby given the number of remaining impressions at the beginning of period t, it sets the fraction α_t such that the optimal expected cost is achieved, assuming that optimal decisions will be made in periods $t+1, ..., T$. Note that, perhaps contrary to one's initial intuition, a greedy policy that assigns high fractions to the advertiser in earlier periods is not necessarily optimal since the supply distributions are time variant. In fact, we can show that the following result is true of any myopic policy (which includes the class of all greedy policies).

Proposition 2.2.1. *Any myopic policy for the single-advertiser ad delivery problem can perform arbitrarily badly compared to the optimal solution.*

A myopic policy by definition does not take the future into account and tries to provide a solution as if the current period is the last or only period in the problem. The following simple example shows that the preceding statement is true.

Example 2.2.2. *An advertiser has a demand of 40 ads, to be delivered over two periods. The cost of overdelivery is* 1 *and the shortage cost is* 3. *In the first period, the supply of ads is a Bernoulli random variable, taking a value of either* 50 *or* 100 *with equal probability. The supply in the second*

period is again a Bernoulli random variable, taking the value 50 *with probability* ϵ *and* 100 *with probability* $1-\epsilon$*. Denote by* α_1 *and* α_2 *the fraction of supply assigned to the advertiser in periods* 1 *and* 2*, respectively, and let the cost of the myopic policy be* $Cost_{myopic}$ *and the cost of the optimal policy be* $Cost_{opt}$*. Any myopic policy will set* $\alpha_1 > 0$ *as it tries to fulfill some of the demand in period* 1*, and therefore incurs positive expected cost. In fact, for this example a myopic policy that tries to optimally balance overdelivery and shortage costs in the first period sets* $\alpha_1 = 0.8$ *and incurs an expected cost of* 20 *in the first period alone. This can be checked as the optimal solution to the following problem*

$$\min_{\alpha_1} E[\alpha_1 X_1 - 40]^+ + 3E[40 - \alpha_1 X_1]^+,$$

which is what the myopic policy tries to solve. As ϵ *goes to zero however, an optimal solution can set* $\alpha_1^* = 0$ *and* $\alpha_2^* = 0.4$*, and the optimal cost approaches zero, making* $\frac{Cost_{myopic}}{Cost_{opt}} \to \infty$*.*

Obviously, as soon as overdelivery occurs in one period and the associated costs are incurred, there is no reason to assign any future supply to the advertiser. One can think of fulfilling the demand over multiple periods as an opportunity to avoid overdelivery in any one particular period by spreading the delivery over the entire horizon.

Unsurprisingly, the sequential nature of the problem lends itself to a dynamic programming framework. Let the state variable at time t be d_t, the number of remaining impressions to be displayed over the rest of the planning horizon. The sequence of events in period t is as follows. d_t is observed and the fraction α_t is set to some value. The supply X_t is then realized and the yield $\alpha_t X_t$ goes towards fulfilling part or all of the advertiser's demand. The state variable for the next period, d_{t+1}, is set equal to $(d_t - \alpha_t X_t)^+$. I will denote by $g_t(d_t)$ the optimal expected cost-to-go function; that is, $g_t(d_t)$ is the optimal expected cost at time t when there are d_t remaining impressions, and assuming that optimal decisions will be made in periods t through T.

2.3 Single Advertiser

I start the analysis by focusing on the case of a single advertiser. It is worth noting that in addition to the benefits of illustrating the structure of the solution in a simplified context, this case is also of relevant practical interest. In the multiple-advertisers case, the advertisers' problems are linked through the constraint that the sum of the fractions of supply assigned to them is at most one. Since in some scenarios it is not uncommon for the publisher to have more supply than the aggregate demand, this constraint becomes non-binding, and the problem can be decoupled into separate single advertiser problems. Taking this view further, I formalize the preceding point in the assumption that follows. Let the optimal fraction in period t, $t = 1, ..., T$ be denoted by α_t^* and consider:

Assumption 2.3.1. *In the optimal solution to the single advertiser delivery problem,* $\alpha_t^* < 1$ *for all* t*.*

As mentioned, one can easily think of scenarios where this assumption would be valid. Indeed, there will be specific target groups and/or various criteria for which it is probably never the case that the publisher assigns all the display opportunities it gets to a single advertiser, since the advertiser's demand is considerably smaller than the available supply, and hence the optimal fraction of ads assigned to that advertiser will always be strictly less than one (as a trivial example, think of an advertiser that wants to display ads to males in the age bracket of 20 to 40 — a very large target audience). On the other hand, one can construct examples where the optimal solution gives the advertiser every single display opportunity that the publisher gets. This may happen if the advertiser is interested in a very unique set of target demographics, such that the supply of the display opportunities for the specified criteria is scarce and barely enough to fulfill the demand. Another possibility is that the cost per undelivered impression is very high compared to the per-impression overdelivery cost, resulting in a very conservative policy that aims to avoid shortage by setting α to its maximum possible value. For the purposes of this section though, and assuming that the above assumption holds, I can derive a simple closed form for the optimal policy that is summarized in the following theorem.

Theorem 2.3.2. *Let d_t be the number of remaining impressions at the beginning of period t. There exist nonnegative numbers $k_1, k_2, ..., k_T$, such that in the ad delivery problem, the optimal policy in period t is to set $\alpha_t^* = d_t/k_t$. Furthermore, computing the values k_t for $t = 1, ...T$ can be done efficiently in an offline (i.e., before the first period begins) manner.*

Proof. I start by solving a single-period problem and then extend the solution to its multi-period counterpart. Consider a single-period problem with demand D and random supply X. A fraction α^* is chosen before X is realized such that α^* is the solution to the following problem

$$\mathcal{L}(D, \alpha) = \min_{\alpha} E[p_1(D - \alpha X)^+ + p_2(\alpha X - D)^+]. \tag{2.1}$$

This expectation can also be written as

$$\mathcal{L}(D, \alpha) = p_1 \int_0^{D/\alpha} (D - \alpha x) \ dF(x) + p_2 \int_{D/\alpha}^{\infty} (\alpha x - D) \ dF(x), \tag{2.2}$$

which can be verified to be a convex function of α. The first derivative of (2.2) with respect to α is

$$\frac{d\mathcal{L}(D, \alpha)}{d\alpha} = -p_1 \int_0^{D/\alpha} x f(x) dx + p_2 \int_{D/\alpha}^{\infty} x f(x) dx. \tag{2.3}$$

Because x is a nonnegative random variable, the integral $\int_a^b x f(x) dx$ is equal to the integral $\int_a^b (1 - F(x)) dx$. The second derivative, again with respect to α, is then equal to

$$\frac{d\mathcal{L}^2(D, \alpha)}{d\alpha^2} = \frac{p_1 D}{\alpha^2} \int_0^{D/\alpha} x f(x) dx + \frac{p_2 D}{\alpha^2} \int_{D/\alpha}^{\infty} x f(x) dx.$$

This expression is greater than zero for any nontrivial specification of the problem (i.e., a specification with $p_1 > 0, p_2 > 0, D > 0$, and a distribution $F(x)$ that does not put all the weight on zero). Hence the function is convex in α and the first-order condition for minimization obtained from setting (2.3)

equal to zero tells us that α^*, the fraction for which the expectation in (2.1) is minimized, satisfies

$$\alpha^* = \sup_{\alpha} \left\{ \frac{\int_0^{D/\alpha} 1 - F(x)dx}{\int_{D/\alpha}^{\infty} 1 - F(x)dx} \right\} \geq \frac{p_2}{p_1}$$

where the inequality, instead of equality, accounts for discrete distributions. Recalling that the integral $\int_a^b (1 - F(x))dx$ for a nonnegative random variable X gives the expectation of X over the interval (a, b), the optimality condition can be interpreted as finding the fraction α^* that divides the support of X into two intervals, $[0, D/\alpha^*]$ and $(D/\alpha^*, \infty)$, such that the ratio of the contribution of these two intervals to the expectation of X is equal to the ratio p_2/p_1. In the case of a discrete distribution, D/α^* would be the first point in the support of X that makes this ratio equal to or bigger than p_2/p_1. If no such point exists, then α^* is set to its maximum value of one, a possibility that I will ignore when I move to the multi-period version under Assumption 2.3.1.

It is not difficult to see that, for values p_1 and p_2 and a certain distribution $F(x)$, there is only one point in the domain of X, call it k, that would satisfy this ratio condition, i.e., there is a unique value k that solves

$$k = \inf_{z} \left\{ \frac{\int_0^z xf(x)dx}{\int_z^{\infty} xf(x)dx} \right\} \geq \frac{p_2}{p_1}. \tag{2.4}$$

Furthermore, computing this point k requires only knowledge of p_1, p_2, and $F(x)$ — it is independent of D and α. This implies that one can pre-compute k before D is known and before the problem commences (i.e., the publisher can compute k offline before the period begins). This value is then used along with the input D to compute $\alpha^* = D/k$. Thus the optimal solution to the one period problem can be written as

$$\alpha^* = \begin{cases} D/k, & 0 \leq D/k < 1; \\ 1, & D/k \geq 1. \end{cases} \tag{2.5}$$

Using Assumption 2.3.1, I write the optimal cost-to-go function as a function of the demand D, substituting the value of α^* from (2.5) into (2.2)

$$g(D) = p_1 \int_0^k d(1 - \frac{x}{k}) \ dF(x) + p_2 \int_k^{\infty} d(\frac{x}{k} - 1) \ dF(x).$$

Note that in this expression, the only variable is d, by rewriting as

$$g(D) = d\left(p_1 \int_0^k (1 - \frac{x}{k})\ dF(x) + p_2 \int_k^\infty (\frac{x}{k} - 1)\ dF(x)\right)$$

we can see that $g(D)$ is a linear function of the form $g(d) = uD$, where u is a nonnegative constant that is equal to $p_1 \int_0^k (1 - \frac{x}{k})\ dF(x) + p_2 \int_k^\infty (\frac{x}{k} - 1)\ dF(x)$.

Having solved the single-period problem, I extend the solution to its multi-period counterpart. Denote the remaining impressions at the beginning of period T by d_T. Since the problem in period T is identical to the single-period problem I just solved, I can find the values k_T and u_T and write $g_T(d_T) = u_T d_T$. Then, moving backwards in time to period $T-1$ and writing the optimal cost-to-go function for that period, I get

$$\begin{aligned} g_{T-1}(d_{T-1}) = \min_{\alpha_{T-1}} E[&p_2(\alpha_{T-1}X_{T-1} - d_{T-1})^+ \\ &+ g_T(d_{T-1} - \alpha_{T-1}X_{T-1})^+]. \end{aligned}$$

Substituting for $g_T(d_T)$ by $u_T d_T$, this expression becomes

$$\begin{aligned} g_{T-1}(d_{T-1}) = \min_{\alpha_{T-1}} E[&p_2(\alpha_{T-1}X_{T-1} - d_{T-1})^+ \\ &+ u_T(d_{T-1} - \alpha_{T-1}X_{T-1})^+] \end{aligned}$$

which is of the same form as (2.1), with p_1 replaced by u_T. I can then solve for the optimal α^*_{T-1} in the exact same way as before, by finding k_{T-1}. The optimal policy in period $T-1$ is then similar to that of a single-period problem: if the number of remaining impressions at the beginning of period $T-1$ is d_{T-1}, then the optimal solution is to set α^*_{T-1} to d_{T-1}/k_{T-1} and the optimal cost-to-go in that period can be written as $g_{T-1}(d_{T-1}) = u_{T-1}d_{T-1}$. Inductively, we deduce that there are values $k_{T-2}, ..., k_1$ which can all be computed in the same way as in the single-period problem, with the optimal policy in any period given as in the statement of the theorem. Thus the problem reduces to solving a sequence of T single-period problems. Again, since the values k_t and u_t depend only on

p_1 p_2, and $F(x_t)$, they can be computed offline.

It remains to show that k_t and u_t can be computed efficiently. Indeed, finding k_t amounts to solving an equation in a single variable in the continuous case and is only slightly more difficult than calculating the expectation of a random variable in the discrete case. For the latter, assume that the maximum number of values the random variable X_t can take is m, and that the probability of $X_t = x$ is given by $p(x)$, then finding k_t involves nothing more than performing binary search on those m values, where at each step of the search the current value m_i is taken as a candidate for k_t and the summation $\sum_{i=0}^{m_i} x_i p(x_i)$ is evaluated and divided by $E(x) - \sum_{i=0}^{m_i} x_i p(x_i)$ and then compared to $\frac{p_2}{p_1}$. A straightforward, naive implementation of this method will take time $O(m \log m)$, which is already fast enough for all practical purposes. Computing u_t takes $O(m)$ time and is dominated by the time it takes to find k_t. Repeating the entire procedure for each period, the overall running time is $O(Tm \log m)$. □

This result makes intuitive sense, and reinforces the discussion after Assumption 2.3.1. For the one-period problem, as k increases with increasing p_2 or decreasing p_1, α^* decreases in order to try and protect against overdelivery, which becomes a more Hcostly penalty. Similarly, imagine that p_1 is very high compared to p_2, then k takes on smaller values, pushing α^* towards one in order to protect against the high cost of underdelivery even when D is not very large.

As an illustration, here is how k can be computed for some well-known distributions.

Uniform: Let $X \sim U[0, 1000]$ and $p_2/p_1 = 0.5$, then the publisher needs to find k that solves

$$\frac{\int_0^k \frac{x}{1000} dx}{\int_k^\infty \frac{x}{1000} dx} = 0.5$$

and k is equal to 577.35. Assume that the ratio p_2/p_1 increases. This indicates that the overdelivery cost increases relative to the shortage cost, in which case one would expect that the publisher will set a lower α to protect against overdelivery. This equates to having higher values for k, which is indeed the case. Let p_2/p_1 be equal to 1.5, then $k = 774.597$. The opposite is of course true when p_2/p_1 decreases. If one sets p_2/p_1 equal to 1/3, then $k = 447.214$.

Exponential: Let $X \sim exp[\lambda]$, $\lambda = 0.001$, and $p_2/p_1 = 0.5$, then the publisher needs to find k that solves

$$\frac{\int_0^k x\lambda e^{-\lambda x}dx}{\int_k^\infty x\lambda e^{-\lambda x}dx} = 0.5$$

The value of k that solves this equation with the above parameters is $k = 1188.834$. Like the previous example, increasing the ratio p_2/p_1 increases k and decreasing it decreases k. The parameter λ is of paramount importance of course. If λ is high, indicating that display opportunities (which are proportional to $\frac{1}{\lambda}$) are few and far between, then k is also very low, implying that α will either be very high or will have to be set equal to 1.

The multi-period solution gives a nice insight into the structure of the problem. The constant u_t for period t can be written as

$$u_t = u_{t+1}\int_0^{k_t}(1 - \frac{x_t}{k_t})dF(x_t) + p_2\int_{k_t}^\infty(\frac{x_t}{k_t} - 1)dF(x_t)$$

where $u_{T+1} = p_1$. From this expression, and depending on the parameters and the distributions in the problem, u_t may or may not be larger than u_{t+1}. One can interpret u_t as the cost of waiting to fulfill an impression in the next period instead of the current period. Sometimes it can be costly to wait, if for example the supply distributions in future periods are not high enough to satisfy the demand, and so u_t is high. Conversely, if it is early in the horizon and future supply distributions look good, then u_t can be low, as it is unlikely that the publisher will be penalized for waiting, and there is little reason to risk overdelivering. Consider the case where supply is IID across periods. Intuitively, one expects that the longer the horizon becomes, the less α_t is for low values of t (i.e. earlier in the horizon). The reason is that there is no reason to risk overdelivery by setting α_t high early on when the horizon is still long as every period has the same supply distribution. This is clear from the following example.

Example 2.3.3. *Consider a 3-period problem where $X \sim U[0, 100]$ in every period, $p_1 = 3$, and $p_2 = 1$. Starting from the last period and solving for k_3, one gets $k_3 = 50$. Evaluating the constant*

u_3

$$\begin{aligned} u_3 &= p_1 \int_0^k (1 - \frac{x}{k}) \ dF(x) + p_2 \int_k^\infty (\frac{x}{k} - 1) \ dF(x) \\ &= 3 \int_0^{50} \frac{1 - \frac{x}{50}}{100} + 1 \int_{50}^{100} \frac{\frac{x}{50} - 1}{100} \\ &= 1. \end{aligned}$$

Using this value for u_3 and solving the problem again for period 2, one gets $k_2 = 71.7107$, from which $u_2 = 0.4142$. Moving to the first period, $k_1 = 84.08$. Thus as we go earlier in the horizon, k_t increases. However, because $d_{t+1} \leq d_t$, it is not necessarily the case that $\alpha_t < \alpha_{t+1}$. Depending on the demand and the realization of X_t, α_t may or may not be less than α_τ for $\tau > t$.

A myopic policy for this example would set $k_t = 50$ for all t. This means that a myopic policy is inclined to deliver more impressions early on in the horizon compared to the optimal policy, and hence runs a higher risk of overdelivering.

Given that the solution can be computed knowing only the costs p_1 and p_2 and the demand distributions, the publisher can use this information about the optimal cost to adjust and negotiate the penalties p_1 and p_2 so that the resulting contract has minimum possible cost given the demands and requirements of the advertiser.

2.4 Single Advertiser — General Case

When Assumption 2.3.1 is violated, the policy in Theorem 2.3.2 is no longer optimal. The reason for this is that the dependence of the optimal cost-to-go on d is not linear but convex, indicating that one needs to evaluate g_t for all values of d_t to successfully apply backwards induction. When this is the case, the problem has a pseudopolynomial time algorithm with a running time $O(TD)$. The direct dependence on D makes the problem intractable. However, one can still develop an ϵ-approximation scheme for the problem. This means that for any given ϵ, we can find a solution that is within ϵ from the optimal solution and that requires time polynomial in the input size and

$1/\epsilon$ to compute. In this section, I prove the following result:

Theorem 2.4.1. *The optimal ad delivery problems admits an ϵ-approximation that can be efficiently computed.*

Consider the general case of the problem when the optimal fraction α_t^*, $t = 1, ..., T$, can take on its maximum value of one. As I will show, this slight change will unfortunately have a strong effect on the complexity of the problem, making it significantly more difficult than the case discussed in the previous section. I start with the following proposition:

Proposition 2.4.2. *The function $g_t(d_t)$ is convex for all t.*

Proof. I prove the proposition by induction. Consider period T, which is equivalent to a single-period problem, as the base case. The optimal α_T^* in this last period is still given by (2.5). If $d_T < k_T$, then the optimal expected cost is convex (in fact, linear) in d_T as shown earlier. If $d_T \geq k_T$ then $\alpha_T^* = 1$, and the optimal expected cost is given by

$$h_T(d_T) = p_1 \int_0^{d_T} (d_T - x_T) dF(x_T) + p_2 \int_{d_T}^{\infty} (x_T - d_T) dF(x_T),$$

which is easily verified to be convex in d_T. This means that $g_T(d_T)$ consists of two parts: a linear function for $d_t < k_t$ and a convex function for $d_T \geq k_T$. For $g_T(d_T)$ to be convex over its entire domain, the slope should be increasing at the break point k_t. To show that this is the case consider the unconstrained problem, where α^* can take on any value regardless of whether $d_T < k_T$ or not, then $g_T(d_T) = u_T d_T$ is a lower bound on the optimal value of $g_T(d_T)$ for all values of d_T. This means that for any value $d_T > k_T$ the graph of the constrained solution can only lie on or above the line $u_T d_T$, which implies a nondecreasing slope at k_T. Another way to see this is to note that the function $\max\{u_T d_T, h_T(d_T)\}$ for values of $d_T \geq k_T$ is convex (the maximum of linear and convex functions), and is always equal to $h_T(d_T)$. It follows that the overall optimal cost function $g_T(d_T)$ is convex on its domain.

For the induction step, assume $g_{t+1}(d_{t+1})$ is convex and write $g_t(d_t) = \min_{\alpha_t} v_t(d_t)$, where

$$v_t(d_t) = \left\{ p_2 \int_{d_t/\alpha_t}^{\infty} (\alpha_t x_t - d_t) dF(x_t) + \int_0^{d_t/\alpha_t} g_{t+1}(d_t - \alpha_t x_t) dF(x_t) \right\}.$$

The first part is convex in α_t and the second is convex by the induction hypothesis and the fact that integration preserves convexity on a monotonically increasing convex function. Thus $v_t(d_t)$ is convex in α_t and can be efficiently minimized. Let the minimizer be α_t^* and write

$$g_t(d_t) = p_2 \int_{d_t/\alpha_t^*}^{\infty} (\alpha_t^* x_t - d_t) dF(x_t) + \int_0^{d_t/\alpha_t^*} g_{t+1}(d_t - \alpha_t^* x_t) dF(x_t).$$

Again, the first part of this expression is a convex function in d_t, while the second part is convex under the induction hypothesis. This finishes the proof of the proposition. □

Solving this problem is equivalent to computing the optimal expected cost at the beginning of the planning horizon when we still have all the demand left to fulfill, i.e., solving the problem is essentially the same as computing $g_1(D)$. As mentioned, the slight change in allowing α^* to be equal to one has a considerable effect on the complexity of the problem. The special case I handled in Section 2.3 involved successively solving a sequence of single-period problems where any two consecutive periods t and $t+1$ are linked together only through a constant u_{t+1} that is easily computed. In the general case however, and because $g_t(d_t)$ is a convex rather than a linear function of d_t for all t, we may have to compute $g_{t+1}(d_{t+1})$ for every value of d_{t+1} in order to be able to compute $g_t(d_t)$. This means that we may have to compute g_{t+1} for every value up to potentially the total demand D. This makes the running time of the problem pseudo-polynomial in the input size, i.e., the complexity of the problem depends directly on the input data instead of its encoding size —which in this case is $\log D$— and the solution exhibits exponential running time behavior (Garey and Johnson (1979)).

To try and alleviate this problem, I will construct a Fully Polynomial Time Approximation Scheme (FPTAS). A minimization problem has an FPTAS if for every $\epsilon > 0$ and for every instance $\mathcal{I}$ of the problem, the algorithm takes time polynomial in the logarithm of the data and in $1/\epsilon$, and

produces a solution $\mathcal{A}(\mathcal{I})$ such that $\mathcal{A}(\mathcal{I}) \leq (1+\epsilon)Opt(\mathcal{I})$,where $Opt(\mathcal{I})$ is the optimal solution to $\mathcal{I}$. The FPTAS I construct for this problem relies on geometric rounding techniques and relies on the fact that the cost-to-go function is monotonic or consists of a bounded number of monotonic functions. Our goal will be to evaluate each g_t at only a subset of values of d_t such that the cardinality of this subset is bounded by a polynomial in the input size, as well as the inverse of the accuracy parameter ϵ. The loss of accuracy is a result of ignoring information by focusing only on a subset of values. The following definitions will be helpful.

Definition 2.4.3. *(δ-approximation function) Let $\delta > 1$ and let $f : D \to \mathbb{R}_+$ be a function. We say that $\hat{f} : D \to \mathbb{R}$ is a δ-approximation of f if for all $d \in D$ we have $f(d) \leq \hat{f}(d) \leq \delta f(d)$.*

Definition 2.4.4. *(δ-approximation set) Let $\delta > 1$ and let $f : [L, U] \to \mathbb{R}_+$ be a monotone function. A δ-approximation set of f is an ordered set $S = \{i_1 < ... < i_r\}$ of integers satisfying*

1. *$L, U \in S \subseteq \{L, ...U\}$;*
2. *for each $j = 1$ to $r-1$, if $i_{j+1} > i_j + 1$, then $\frac{f(i_j)}{\delta} \leq f(i_{j+1}) \leq \delta f(i_j)$.*

Let $f : [L, U] \to \mathbb{R}_+$ be a monotonically increasing function with maximum value f^{max}. Let t_f be the time it takes to evaluate f. A δ-approximation set of f can be computed in time $O(t_f \log_\delta f^{max} \log(U - L))$ by performing binary search on $[L, U]$. A δ-approximation function is constructed from a δ-approximation set using the following definition.

Definition 2.4.5. *Let $\delta > 1$ and let $f : [L, U] \to \mathbb{R}_+$ be a monotonically increasing function. Let S be a δ-approximation of f. A function $\hat{f}$ defined as follows is called the approximation of f corresponding to S: For any x such that $L \leq x \leq U$ and successive elements $i_k, i_{k+1} \in S$ with $i_k < x \leq i_{k+1}$, we set $\hat{f}(x) = f(i_{k+1})$.*

I now proceed with approximating the problem. Consider the last period. As I have shown, calculating the value k_T for that period is not difficult, but I need to calculate the value of $g_T(d_T)$ for each value of d_T whenever $d_T > k_T$. Depending on the distribution and the costs p_1 and p_2, k_T might be quite low, and I would have to calculate d_T for a number of values that is effectively on the

order of the total demand D. This motivates us to use the previous definitions to limit our attention to only a subset of values of d_T, namely, the δ-approximation set of $g_T(d_T)$. Because $g_T(d_T)$ is a convex, monotonically increasing function, I can indeed construct a δ-approximation set for it and then use Definition 2.4.5 to construct $\hat{g}_T(d_T)$, a δ-approximation function of $g_T(d_T)$. The following lemma follows immediately from Definition 2.4.3.

Lemma 2.4.6. *For any value of d_T in the domain of the last period, $g_T(d_T) \leq \hat{g}_T(d_T) \leq \delta g_T(d_T)$.*

Now that I have an approximation of the value function in the last period, I can move backwards in time to approximate $g_{T-1}(d_{T-1})$. I will drop the subscript t when I talk about the demand from now on, using $g_{T-1}(d)$ instead of $g_{T-1}(d_{T-1})$. One problem is that $g_T(d)$ is used to calculate $g_{T-1}(d)$ and we have no access to $g_T(d)$, but instead have its approximation $\hat{g}_T(d)$. One can intuitively see that using $\hat{g}_T(d)$ in place of $g_T(d)$ while evaluating $g_{T-1}(d)$ will result in an error in the value of $g_{T-1}(d)$, and as we repeat the process and approximate $g_{T-1}(d)$ and use its approximation to calculate $g_{T-2}(d)$, the error gets worse. This is to be expected, as we are using an approximate function as part of another function we are approximating, so the error gets compounded. Before examining this, I write the minimization problem for a fixed d in the penultimate period, namely how to find α^*_{T-1} that solves

$$g_{T-1}(d) = \min_{\alpha_{T-1}} E\left[p_2 \left(\alpha_{T-1}x_{T-1} - d\right)^+ + g_T(d - \alpha_{T-1}x_{T-1})^+\right]. \tag{2.6}$$

Because the second term in the expectation is not available, I use its approximation, $\hat{g}_T(d)$, instead. We are then looking for the value of α_{T-1} that minimizes

$$p_2 \int_{\frac{d}{\alpha_{T-1}}}^{\infty} (\alpha_{T-1}x_{T-1} - d)dF(x_{T-1}) + \int_0^{\frac{d}{\alpha_{T-1}}} \hat{g}_T(d - \alpha_{T-1}x_{T-1})dF(x_{T-1}). \tag{2.7}$$

Because of the $\hat{g}_T$ term, this function is not necessarily continuous. I will define $\bar{g}_T$ as the piecewise linear extension of $\hat{g}_T$, so that $\bar{g}_T$ is both continuous and convex. The piecewise linear extension of a function f on a subset $S = [a, b]$, where a and b are integers, is the continuous function obtained by making f linear between successive values of S. Convexity of $\bar{g}_T$ follows from the fact

that the points in $\hat{g}_T$ come from the convex function g_T. I then use $\bar{g}_T$ in place of $\hat{g}_T$ in (2.7) above to define

$$\check{g}_{T-1}(d) = \min_{\alpha_{T-1}} E\left[p_2\left(\alpha_{T-1}x_{T-1} - d\right)^+ + \bar{g}_T(d - \alpha_{T-1}x_{T-1})^+\right]. \tag{2.8}$$

This is a convex minimization problem that can be solved efficiently and whose minimizer I will denote by $\hat{\alpha}_{T-1}$. I would like to understand how the solution produced by $\hat{\alpha}_{T-1}$ on $\check{g}_{T-1}(d)$ compares to the solution produced by the optimal α^*_{T-1} on the original problem of minimizing $g_{T-1}(d)$. The relationship is summarized in the following simple lemma.

Lemma 2.4.7. *For any $0 \leq d \leq D$, we have $\check{g}_{T-1}(d) \leq \delta g_{T-1}(d)$.*

Proof. From Lemma 2.4.6 and Definition 2.4.5, we know that for any value of d in $[a, b]$, where a and b are in the δ-approximation set of g_T, we have $\hat{g}_T(d) = \hat{g}_T(b) \leq \delta g_T(d)$. Since $\bar{g}_T$ is linear between any two consecutive points in the δ-approximation set (like a and b here), the relationship $\bar{g}_T(d) \leq \hat{g}_T(b)$ holds, and therefore $\bar{g}_T(d) \leq \delta g_T(d)$. Comparing equations (2.6) and (2.8), we see that the first term in both expectations is the same, and for any value of α_{T-1} we have $\bar{g}_T(d - \alpha_{T-1}x_{T-1}) \leq \delta g_T(d - \alpha_{T-1}x_{T-1})$ as shown. Consider α^*_{T-1} as a solution to (2.8). By the preceding discussion, the value produced by this solution is such that $\check{g}_{T-1}(d) \leq \delta g_{T-1}(d)$. It follows that there exists a minimizer $\hat{\alpha}$ such that the relationship given in the statement of the lemma holds. □

I have thus shown that for a fixed value of d, we can find a solution to the penultimate period that is not more than a multiplicative error of δ away from the optimal solution for this value of d in that period. I then proceed to find $\hat{g}_{T-1}(d)$, the delta approximation function of $\check{g}_{T-1}(d)$, as before. Notice that as we do so, we accumulate more errors since $\hat{g}_{T-1}(d) \leq \delta\check{g}_{T-1}(d)$ by Definition 2.4.3, and hence $\hat{g}_{T-1}(d) \leq \delta^2 g_{T-1}(d)$ by Lemma 2.4.7. The whole process is repeated for each of the time periods $T-2, ..., 1$. The following lemma generalizes Lemma 2.4.7 and summarizes the relationship between $\hat{g}_t(d)$ and $g_t(d)$ for all t.

Lemma 2.4.8. *In period t, $t = 1, ..., T$, we have $\hat{g}_t(d) \leq \delta^{T+1-t} g_t(d)$.*

Proof. I prove the lemma by induction on t. From Corollary 2.4.6, we know that the result holds for the base case $t = T$. The proof for period t is similar to the arguments I considered for

period $T-1$. Assume inductively that the relationship holds for period $t+1$ and consider $\check{g}_t(d) = \min_{\alpha_t} E\left[p_2\left(\alpha_t x_t - d\right)^+ + \bar{g}_{t+1}(d) - \alpha_t x_t)^+\right]$. The first term in this expectation is the same as that in the problem of minimizing $g_t(d)$, and by the induction hypothesis $\hat{g}_{t+1}(d) \leq \delta^{T-t} g_{t+1}(d)$ and hence $\bar{g}_{t+1}(d) \leq \hat{g}_{t+1}(d) \leq \delta^{T-t} g_{t+1}(d)$. Therefore by the same arguments as in Lemma 2.4.7, there exists a minimizer for $\check{g}_t(d)$ such that $\check{g}_t(d) \leq \delta^{T-t} g_t(d)$. Calculating the δ-approximation function $\hat{g}_t(d)$ for $\check{g}_t(d)$ and using Definition 2.4.3, we have $\hat{g}_t(d) \leq \delta \check{g}_t(d) \leq \delta \delta^{T-t} g_t(d) = \delta^{T+1-t} g_t(d)$. □

With Lemma 2.4.8 in place, I can give the main result of this section.

Theorem 2.4.9. *For any $\epsilon \in (0,1]$, the ad delivery problem admits an FPTAS by setting $\delta = 1 + \frac{\epsilon}{2T}$. That is, we can find $\hat{g}_1(D)$ such that $g_1(D) \leq \hat{g}_1(D) \leq (1+\epsilon) g_1(D)$ by using this value of δ to approximate $g_t, t = 1, ..., T$.*

Proof. From Lemma 2.4.8, we have $\hat{g}_1(D) \leq \delta^T g_1(D)$. Setting δ equal to the value in the statement of the theorem we have $\hat{g}_1(D) \leq (1 + \frac{\epsilon}{2T})^T g_1(D)$. Because $1 + \frac{\epsilon}{2T} = 1 + \frac{\epsilon/2}{T}$, we can use the inequality $(1 + \frac{x}{n})^n \leq 1 + 2x$ which holds for every $x \in [0,1]$ to get $\hat{g}_1(D) \leq (1+\epsilon) g_1(D)$. It remains to show that the time taken by the algorithm is polynomial in the input size. Consider one iteration of the algorithm and let the largest value produced by the algorithm at any stage be B. Because we know that the values produced by the algorithm are at most δ^T away from the values produced by the optimal algorithm, the upper bound B is polynomial in the size of the problem and in δ (or, equivalently, $1/\epsilon$). Evaluating $\bar{g}_t$ takes time t_g and finding a δ-approximation set S takes time $O(t_g \log_\delta B \log D)$. Because $0 < \epsilon \leq 1$, we know $\delta < 2$, and using the relationship given in the statement of the theorem, I can rewrite the time it takes to compute a δ-approximation set as $O(t_g \frac{T}{\epsilon} \log B \log D)$. Finding a convex extension for $\hat{g}_t$ is linear in the size of S, and is dominated by the time it takes to compute the δ-approximation set. Repeating these steps for each period, the overall running time is given by $O(t_g \frac{T^2}{\epsilon} \log B \log D)$, which is polynomial in the input size, the number of periods, and $1/\epsilon$, as desired. □

Notice that, because of the way the approximation works, there is a tendency to overdeliver impressions, but not by much. This happens because, by (2.4.5), for any value of demand $d \in [a,b]$

at the beginning of a period where a and b are in the δ-approximation set of the value function in that period, the algorithm operates as if the remaining number of impressions is $b \geq d$. Nevertheless, we are assured that in doing so the extra expected cost at the end of the horizon will not be more than a multiplicative factor of ϵ away from the optimal solution.

2.5 Extensions

2.5.1 Multiple Advertisers

I will slightly revise the structure of the costs before I extend the results to multiple advertisers. Throughout the preceding discussion, I have interpreted the penalty p_2 as the opportunity cost of giving away an impression for free instead of selling it to another advertiser. When I consider the multiple advertisers case under this interpretation, there is no reason to keep the overdelivery costs, since the case where advertiser i is allocated more impressions than their demand only impacts the solution if this overdelivery results in shortage for other advertisers, and hence the penalty p_2 can be implicitly incorporated into the shortage costs of advertisers other than i. I will discuss the case when the advertiser also wishes to not receive extra impressions over their demand in the next subsection. For this section, I assume that there are m advertisers and that advertiser i's shortage cost is given by p_i. The decision vector in period t is $\alpha^{\mathbf{t}} = (\alpha_1^t, ..., \alpha_m^t)$, where α_i^t is the fraction assigned to advertiser i in period t. The problem then is the same as before: the publisher is interested in choosing $\alpha^{\mathbf{t}}, t = 1, ..., T$ in order to minimize the shortage costs at the end of the horizon. One difference is that all cost is evaluated at the last period, since there is no longer an overdelivery cost in any one period. Formally, I solve $\mathcal{L}(\mathbf{d}, \alpha)$, where $\mathbf{d}$ is the vector of demands, and

$$\begin{aligned} \mathcal{L}(\mathbf{d}, \alpha) = \min_{\alpha^{\mathbf{t}}} \quad & E\left[\sum_i p_i \left(d_i - \sum_t \alpha_i^t x^t\right)^+\right] \\ \text{s.t.} \quad & \sum_i \alpha_i^t \leq 1 \quad t = 1, ..., T \\ & 0 \leq \alpha_i^t \leq 1 \quad t = 1, ..., T.\ i = 1, ..., m \end{aligned}$$

I assume that the constraints $\sum_i \alpha_i^t \leq 1$ are binding. This is without loss of generality, since

one can introduce a dummy advertiser that gets assigned any leftover impressions in a period if the constraint has some slack. Furthermore, if it is the case in the optimal solution that $\sum_i \alpha_i^{*^t} < 1$ for all t, then the problem can simply be decoupled into m separate problems that are then solved as in the previous section. Starting again from the one-period problem, one can verify convexity in α as in the single-advertiser case. The constraints are linear in $\alpha_1, ..., \alpha_m$ and the Hessian matrix of the objective is positive definite.

Under the binding constraints assumption, advertiser m is assigned a fraction $1 - \sum_{i \neq m} \alpha_i$, so that setting the fractions for all but the last advertiser automatically determines the fraction that the last advertiser gets. Rewriting the single-period objective in the form of (2.2), I get

$$\begin{aligned} \mathcal{L}(\mathbf{d}, \alpha) = p_1 \int_0^{d_1/\alpha_1} (d_1 - \alpha_1 x) \ dF(x) + ... \\ + p_m \int_0^{\frac{d_m}{1-\sum_{i \neq m} \alpha_i}} (d_m - (1 - \sum_{i \neq m} \alpha_i)x) \ dF(x). \end{aligned} \tag{2.9}$$

Notice that when minimizing (2.9), I end up with a system of $m-1$ equations, corresponding to the $m-1$ decision variables $\alpha_1, ...\alpha_{m-1}$. Since each equation is the partial derivative of (2.9) with respect to one of the variables, it has exactly two terms: the derivative of the integral that contains that variable as well as the derivative of the last integral, which is expressed in terms of the first $m-1$ variables. Specifically, the derivative of (2.9) with respect to α_i is given by

$$\frac{d\mathcal{L}(\mathbf{d}, \alpha)}{d\alpha_i} = -p_i \int_0^{d_i/\alpha_i} x f(x) + p_m \int_0^{\frac{d_i}{1-\sum_{i \neq m} \alpha_i}} x f(x).$$

Note that in particular, the second term is common to all equations. Writing this out for all the $m-1$ variables and equating each derivative to zero to obtain the conditions for minimization, I find that for any two advertisers i and j, the following holds at the optimal solution

$$p_i \int_0^{d_i/\alpha_i} x f(x) = p_j \int_0^{d_j/\alpha_j} x f(x). \tag{2.10}$$

Like before, we will let $k_i = \frac{d_i}{\alpha_i}$. The optimal solution to the problem then involves finding

$k_1, ..., k_{m-1}$ such that Condition (2.10) is satisfied for all i and j. In addition, since determining $k_i, i = 1, ..., m-1$ determines $\alpha_i, i = 1, ..., m-1$, it also determines α_m through the relation $\alpha_m = 1 - \sum_{i \neq m} \alpha_i$. The resulting α_m should satisfy Condition (2.10). Without loss of generality, let the costs p_i be arranged such that $p_1 \geq p_2 \geq ... \geq p_m$. If we follow the approach from the previous section, we can try to find values of k_i such that the following holds for all i and j

$$\frac{\int_0^{k_i} xf(x)}{\int_0^{k_j} xf(x)} = \frac{p_j}{p_i}. \tag{2.11}$$

A set of values for $k_i, i = 1, ...m$ that solves (2.11) and leads to a vector α with $\sum_i \alpha_i = 1$ gives a solution to the problem. From (2.11) and the fact that X is a nonnegative random variable, one can see that advertisers with low index have lower k values. The immediate implication is that these advertisers get more share of the supply if the demands of all advertisers are the same or comparable (since low k values correspond to high values for α when the demands are the same). This agrees with intuition and suggests that the optimal single-period policy has a greedy flavor, allocating more shares to those advertisers that have higher penalties. In fact, it is possible that advertisers with high indices (low p_i) get assigned zero impressions, since the only way the condition is satisfied is if their corresponding values of k_i are set to infinity. Of course, since the conditions above also depend on d_i, it is not always the case that high-index advertisers receive fewer impressions — the important thing is that the optimality conditions are satisfied.

When one considers the multiple-period problem, applying the same policy in a myopic fashion turns out to again be suboptimal. Consider the following example:

Example 2.5.1. *Assume there are two advertisers with demands* $d_1 = 30$ *and* $d_2 = 60$ *and* $p_1 = 20$, $p_2 = 1$. *There are two periods, with* X_1 *in the first period being distributed uniformly over* $[0, 100]$ *and in the second-period an almost degenerate distribution on 30, so that* $Pr(X_2 = 30) = 1 - \epsilon$. *A myopic policy assigns* $\alpha_1^1 = 1$ *and* $\alpha_2^1 = 0$. *Thus whatever happens in the first period, the second advertiser will get at most 30 impressions, and the cost is bounded below by* $30p_2 = 30$. *Consider a policy that instead sets* $\alpha_1^1 = 0, \alpha_2^1 = 1, \alpha_1^2 = 1$, *and* $\alpha_2^2 = 0$, *so that it fulfills all of the first advertiser's demand*

in the second period and gives all the yield from the first period to the second advertiser, then this policy only incurs the cost of the unfulfilled impressions that the second advertiser does not get. Thus as $\epsilon \to 0$, the cost is given by

$$\int_0^{60} \frac{60 - \alpha_2^1 x_1}{100} dx = 18,$$

which is less than the cost of the myopic policy.

2.5.2 Additional Delivery Constraints

Let us return to the single-advertiser case. So far, the publisher's problem has been of the form

$$\min_{0 \le \alpha_t \le 1} \sum_t h(d_t, F_t(x), \alpha_t)$$

with $h(d, F(x), \alpha)$ taking the form of the function in (2.1). I want to consider allowing the advertiser to have more input into the structure of the delivery process, specifically, the advertiser can choose a function $l(d_t, F_t(x), \alpha_t)$ such that the publisher's objective becomes

$$\min_{0 \le \alpha_t \le 1} \sum_t h(d_t, F_t(x), \alpha_t) + l(d_t, F_t(x), \alpha_t).$$

I illustrate this in the context of the discussion at the beginning of this chapter, where in addition to the guaranteed delivery requirement, the advertiser would like its ads to be evenly spaced over time. An advertiser with total demand D over a horizon of length T can then choose $l(d_t, F_t(x), \alpha_t) = q|\alpha_t x_t - \frac{D}{T}|$, so that there is a penalty q associated with delivering more or less than D/T impressions in each period (of course, the advertiser can specify any other value than D/T, or different values for different periods). For simplicity, let us roll the costs p_1 and p_2 into a single cost p. The publisher's problem then becomes

$$\min_{0 \le \alpha_t \le 1} E\left[p \left| D - \sum_{t=1}^{T} \alpha_t X_t \right| + q \left| \sum_{t=1}^{T} \alpha_t X_t - \frac{D}{T} \right| \right].$$

This problem closely follows the framework outlined above, both for the special case under Assump-

tion 2.3.1 and the general case (depending on the relationship between p and q, it may be necessary to set α equal to 1 in some scenarios). Just to illustrate, under Assumption 2.3.1 the optimal α in a single-period problem satisfies

$$\alpha^* = \sup_{\alpha} \frac{\int_{\frac{D\alpha}{T}}^{\infty} xf(x)dx - \int_0^{\frac{D\alpha}{T}} xf(x)dx}{\int_0^{\frac{D}{\alpha}} xf(x)dx - \int_{\frac{D}{\alpha}}^{\infty} xf(x)dx}.$$

As one can tell from this expression, the criteria for optimality looks more complex as one adds more requirements. Nevertheless, the structure of the solution (finding intervals that divide the domain of the distribution in a certain way) remains the same. It turns out that a sufficient condition to add more expressiveness while maintaining the general flavor of the solution is the requirement that $l(d_t, F_t(x), \alpha_t)$ be convex, which makes the publisher's overall objective convex in α_t and d_t. If $l(d_t, F_t(x), \alpha_t)$ is chosen such that, for example, there is a bonus paid to the publisher once a certain target $z < D$ is fulfilled, then the objective displays a kink and convexity is destroyed. In such scenario, the methods outlined in this chapter may fail to be optimal.

2.6 Discussion

This chapter provides optimal policies to some variants of the guaranteed delivery problem in display advertising. I have shown that when the advertiser's demand is low compared to the overall supply, the problem can be solved to optimality and the optimal policy has a nice and simple characterization. Because the publisher is able to calculate its expected cost as a function of the demand D and costs p_1 and p_2, it can use this information in deciding on prices to charge the advertiser for service, as well as negotiate the shortage penalty p_1. For the general case, the dynamic program becomes computationally difficult to solve and I provide an approximation scheme that gives a trade-off between the complexity of the problem and the quality of solution produced.

The case for multiple advertisers maintains the same spirit of the solution, namely, dividing the support of the distribution into intervals from which the optimal fractions can be calculated. Figuring out the fractions for the single-period multiple-advertisers case is not as straightforward

as the single-advertiser one. If instead of the modification we introduced in the multiple-advertiser scenario we had each advertiser still maintain under- and overdelivery penalties then a myopic policy is no longer optimal and the problem becomes quite difficult even to approximate.

There are many variations on the theme of this problem. I have discussed a sufficient condition under which the methods presented here extend to more expressive contracts, namely, the convexity of the publisher's objective function. It would be interesting to identify the correspondence between bids and fractions: we know what fraction the publisher should set in the optimal solution to the problem, but in reality, and as mentioned in the introduction, the publisher places a bid in an auction for a period of time, not a fraction. The interaction between maximum prices that advertisers are willing to pay per impression and the bids placed by the publisher affects the fractions that the advertiser can select and therefore the structure of the optimal delivery policies. It would therefore be instructive to understand how the two separate processes of selecting optimal contracts and fulfilling these contracts interact, so that instead of designing each in isolation one can develop a more integrated approach that accounts for the issues addressed by each.

Chapter 3

Exploiting Myopic Learning

3.1 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 *non-Bayesian* 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 gains. In this chapter, 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

[1] For 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).

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 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 chapter 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's 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 3.2.1, 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 are 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

[2]See, for example, Bomze (1986), Fischer and Vocking (2004), Fischer, Racke, and Vocking (2006), and the survey in Fudenberg and Levine (1998).

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 Nash outcome if the population learns according to replicator dynamics.

The primary contribution of this chapter 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 chapter can be summarized as follows:

1. If the principal is myopic, reacting 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 chapter 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. I provide a real-world example that shows that such optimal policies are possibly already implemented in practice in certain contexts.

This chapter is related to several strands of literature on (behavioral) mechanism design and

[3] One 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.

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 chapter 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 become very large, 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 chapter 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 chapter is organized as follows. Section 4.2 presents the class of games that I will focus on for the rest of the chapter. Section 3.3 discusses the case of a myopic population *and* a myopic principal. Section 3.4 derives the optimal policies when the principal is forward-looking and Section 3.5 discusses how these policies improve over the rational population case. Section 3.6 provides some empirical examples that support the predictions of the model. Section 3.7 gives other applications for the methods used in the chapter. Section 3.8 quantifies how the degree of sophistication of the population impacts the principal, and Section 3.9 concludes the chapter.

3.2 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

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

Figure 3.1: The Cheat-Audit Game

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.

3.2.1 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 3.1, 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, 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 3.2.1. *(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 3.1 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

$$\begin{aligned} \alpha_N &= \frac{v_3 - v_2}{v_3}; \\ x_N &= \frac{c_2}{c_3 + c_2 - c_1}. \end{aligned} \tag{3.1}$$

As mentioned, I consider an infinitely repeated setting where at each moment in time the game in Figure 3.1 is played. Discrete time and how it affects the results I obtain is discussed in Section 3.6. I will let the state of the system at time t be the fraction of the population taking action C

[4] One 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 having been caught and does not affect the probability of being detected.

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.

Proposition 3.2.2. *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 Proposition 3.2.2 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.

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

$$\begin{aligned} g(x(t),\alpha(t)) &= c_1\alpha(t)x(t) + c_2\alpha(t)(1-x(t)) + c_3(1-\alpha(t))x(t) \\ &= (c_1 - c_2 - c_3)\alpha(t)x(t) + c_2\alpha(t) + c_3x(t), \end{aligned} \tag{3.2}$$

where the terms in the first equation in (3.2) 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.

3.2.2 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 3.2.3. *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 3.1, 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 the quantities above. 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 $\alpha(t)$

$$\begin{aligned} \dot{x}(t) = f(x(t), \alpha(t)) &= q(1 - \alpha(t))x(t)(1 - x(t)) - p\alpha(t)x(t)(1 - x(t)) \\ &= x(t)(1 - x(t))(q - \alpha(t)(q + p)). \end{aligned} \tag{3.3}$$

3.3 Myopic Principal

Before discussing the optimal policy for the principal, I consider the following question: what happens if the principal 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 learns in a myopic fashion and 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 3.2.3.

Assumption 3.3.1. *At the beginning of the horizon the princpal assigns positive weights to each strategy, 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 they are played 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 (3.1) 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 3.3.2. *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* (3.1) *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 (3.1), 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 3.2 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 that following system

$$\alpha(t)(1-\alpha(t))[(c_3+c_2-c_1)x(t)-c_2) \tag{3.4}$$

and

$$x(t)(1-x(t))[(v_2-v_3-v_1)\alpha(t)+v_3-v_2] \tag{3.5}$$

As Theorem 3.3.2 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 3.6, this cyclical nature can be observed in various real-world phenomena that correspond to the Cheat-Audit game.

3.3.1 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 (3.1). The following result formalizes this fact.

Theorem 3.3.3. *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 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.

3.4 Forward-Looking Principal

A forward looking principal differs from the myopic principal of the last section in taking the future into account, so that instead of reacting to the latest round of the game, the principal 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.

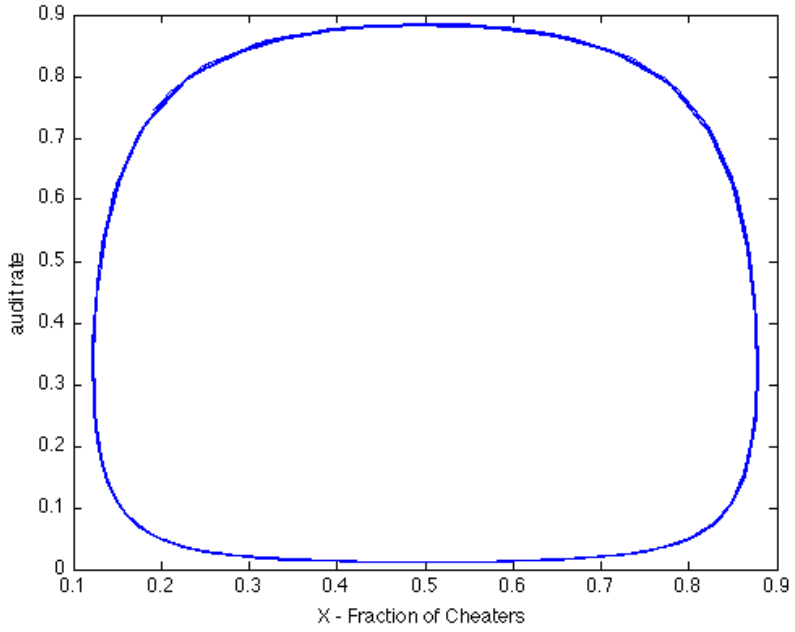


Figure 3.2: Phase Portrait of Cheating and Auditing Activity

3.4.1 Objective

The principal's problem is the following: Given the different values in Figure 3.1 and the learning dynamics, the principal is interested in minimizing the 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 (3.2) and the equation of motion of the population by (3.3). The principal's problem can then be written as

$$\begin{aligned} \min_{\alpha(t)} \quad & \int_0^\infty e^{-rt}((c_1 - c_2 - c_3)\alpha(t)x(t) + c_2\alpha(t) + c_3x(t))dt \\ \text{s.t.} \quad & \dot{x}(t) = x(t)(1 - x(t))(q - \alpha(t)(q + p)) \\ & 0 \leq \alpha(t) \leq 1 \end{aligned} \tag{3.6}$$

where $r > 0$ is a discount factor. Thus the principal's problem involves finding the function $\alpha^*(t)$ that solves (3.6). 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 $\alpha(t)$.

3.4.2 Optimal Policy

3.4.2.1 Single Round

Before trying to find the optimal solution to (3.6), I first develop an intuition by considering the solution when the game is played only once. The stage game cost described by (3.2) 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

$$\alpha^* = \begin{cases} 0, & x < \frac{c_2}{c_2+c_3-c_1}; \\ 1, & x \geq \frac{c_2}{c_2+c_3-c_1}. \end{cases} \tag{3.7}$$

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:

$$g^*(x) = \begin{cases} c_3 x, & x < \frac{c_2}{c_2+c_3-c_1}; \\ c_2 + (c_1 - c_2)x, & x \geq \frac{c_2}{c_2+c_3-c_1}. \end{cases} \tag{3.8}$$

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.

3.4.2.2 General Policy

I will derive the optimal policy for (3.6) 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

$$\begin{aligned} H(x,\alpha,\lambda) &= g(x,\alpha) + \lambda f(x,\alpha) \\ &= c_3 x + \alpha(c_2 + (c_1 - c_2 - c_3)x) + \lambda x(1-x)(q - \alpha(q+p)) \\ &= c_3 x + \lambda q x(1-x) + \alpha(c_2 + (c_1 - c_2 - c_3)x - \lambda(p+q)x(1-x)) \end{aligned} \tag{3.9}$$

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 3.4.1. *The optimal control for Problem (3.6) is a bang-bang solution, where*

$$\alpha^*(t) = \begin{cases} 0, & \lambda(t) < \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}; \\ 1, & \lambda(t) > \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}; \\ [0,1], & \lambda(t) = \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}. \end{cases} \tag{3.10}$$

Proof. See Appendix. □

Lemma 3.4.1 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 (3.6) as a calculus of variations problem. From (3.3), I can express $\alpha(t)$ as

$$\alpha(t) = \frac{1}{p+q}\left(p - \frac{\dot{x}(t)}{x(t)(1-x(t))}\right). \tag{3.11}$$

Substituting this into the objective, the problem becomes

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

$$= \min_{x(t)} \int_0^\infty e^{-rt}\left(c_3 x(t) + \frac{(c_2 + (c_1 - c_2 - c_3)x(t))\left(p - \frac{\dot{x}(t)}{(1-x(t))x(t)}\right)}{p+q}\right) dt. \tag{3.12}$$

The solution to (3.12) provides another necessary condition on the optimal state trajectory. Specifically, the following lemma shows that there is a constant for which the integral in (3.12) is stationary, i.e., the function that minimizes (3.12) is time-independent.

Lemma 3.4.2. *Let $x^*(t)$ be the minimizer to (3.12), then $x^*(t) = C$, where C is a constant that depends on the parameters of the problem and is equal to*

$$\frac{(c_2 - c_1)p - c_3 q + (c_1 - c_2 - c_3)r + \sqrt{4c_2((c_2 - c_1)p - c_3 q)r + ((c_1 - c_2)p + c_3 q + (c_1 r - c_2 - c_3)r)^2}}{2((c_2 - c_1)p - c_3 q)}. \tag{3.13}$$

Proof. See Appendix. □

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

[5] As I show in the proof of Theorem 3.4.3, the third case happens only for precisely a single pair $(x^*(t), \alpha^*(t))$ in the optimal solution.

Theorem 3.4.3. *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 3.4.3 indicates that, depending on a threshold value, the optimal solution either audits indiscriminately or does nothing. If the system hits the value $|barx$, 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 3.6, the optimal policy can oscillate too, as a result of considering the model in discrete time.

3.5 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 3.2.1 and in Theorem 3.2.2 that the (fully rational) repeated game possesses a unique equilibrium in mixed strategies, given by (3.1). As I have shown in Theorem 3.3.2, 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 3.4 is not the Nash equilibrium, indicating that the Nash solution is dominated by the policy in Theorem 3.4.3. 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 3.4.3, 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 concerned enough about the future, keep the fraction of C players close to zero.

Theorem 3.5.1. *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 $\lim_{r\to 0} \bar{x} = 0$, 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 3.2.2 shows, the principal cannot influence the future actions of the population. Furthermore, from Theorem 3.3.3, 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 3.3.2 and 3.4.3 are widely observed in practice.

3.6 Examples

Eeckhout, Persico, and Todd (2010) define crackdowns as intermittent periods of high-intensity monitoring. Crackdown cycles occur when these periods are interwoven 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 crack 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 the Chinese government's methods for controlling 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 3.2 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 China[6], 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 3.3 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, 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 3.4.3, where a severe crackdown brings the fraction of offenders down to a certain level, after which auditing

[6]Lui (1986)
[7]Source: Administrative Office Of The Courts
[8]Source: PEW Internet and American Life Project Data Memo
[9]However, the *amount* of copyrighted material shared online continues to grow.

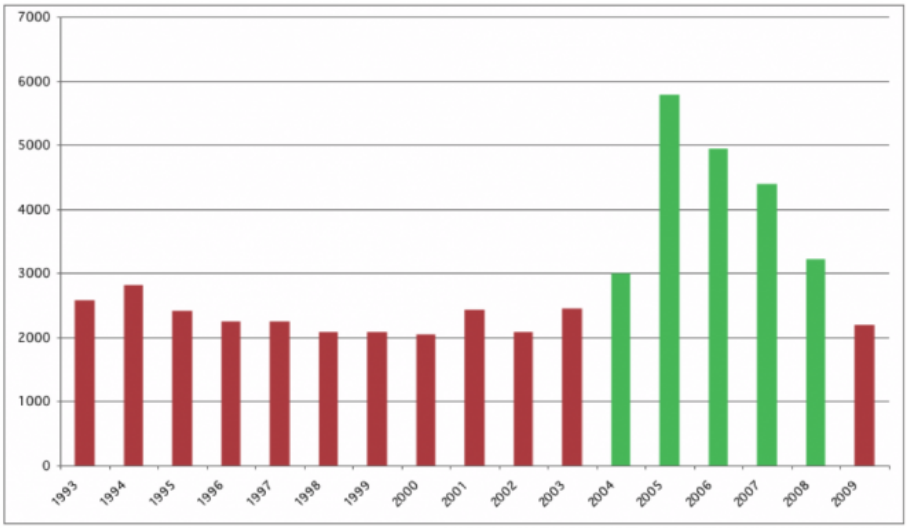


Figure 3.3: US Copyright Infringement Cases 1993-2009

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 chapter suggests that the core driving factors are captured by the model.

The analysis in this chapter 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 chapter. 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.

3.7 Other Applications

3.7.1 Equilibrium Selection and Technology Adoption

The framework I use in this chapter 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 3.4, 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, 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 make 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.

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

Figure 3.4: Coordination Game

The situation described above is depicted in the coordination game in Figure 3.4, 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 4.2, 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 techology L and $1 - x(t)$ as the fraction sticking with the old technology. By Assumption 3.2.3, 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 3.4 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 3.7.1. *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 3.4.

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 4.2 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 (3.6), the principal wants to solve

$$\begin{aligned} \min_{\alpha(t)} \quad & \int_0^\infty e^{-rt}(\alpha(t)(1 - x(t)) + (1 - \alpha(t))x(t))dt \\ \text{s.t.} \quad & \dot{x}(t) = x(t)(1 - x(t))\big((1 - \alpha(t))a_2 - \alpha(t)a_1\big) \\ & 0 \leq \alpha(t) \leq 1. \end{aligned} \tag{3.14}$$

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 (3.7), where

$$\alpha^* = \begin{cases} 0, & x < \frac{b_1}{b_1 + b_2}; \\ 1, & x \geq \frac{b_1}{b_1 + b_2}. \end{cases} \tag{3.15}$$

In the same way as in Lemma 3.4.1 and Theorem 3.4.3, Problem (3.14) 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 3.7.2. *Under Assumption 3.2.3 and as $r \to 0$, a forward-looking principal has a strategy that converges to the (T, L) equilibrium of the coordination game in Figure 3.4.*

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, it 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.

3.8 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 have so far considered in this chapter. 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

[10] In 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.

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 3.2.2 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.

3.8.1 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 3.8.1. *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*

$$p_c^*(x) = \begin{cases} 0, & x > x_N; \\ 1, & x + \rho < x_N; \\ \frac{x_N - x}{\rho}, & x + \rho \geq x_N. \end{cases} \tag{3.16}$$

where x_N is as given in (3.1).

Proof. See Appendix. □

The reason the rational agents behave in this manner can be inferred from the principal's optimal strategy in Section 3.4.2.1. 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 its 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.

3.8.2 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 3.8.1 as $p_c^*(\alpha(t))$ Formally, the rate of change $\dot{x}(t)$ is given by

$$\dot{x}(t) = (1 - x(t) - \rho)(x(t) + p_c^*(\alpha(t))\rho)(1 - \alpha(t))q - \alpha(t)x(t)(1 - x(t) + (1 - p_c^*(\alpha(t)))\rho)p \quad (3.17)$$

where q and p are as before. The first term is the fraction of honest agents that switch 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 3.4.2.2. 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

$$\begin{aligned}\min_{\alpha(t)} \quad & \int_0^\infty e^{-rt}(c_1 - c_2 - c_3)\alpha(t)(x(t) + p_c^*(\alpha(t))\rho) + c_2\alpha(t) + c_3(x(t) + p_c^*(\alpha(t))\rho)\, dt \\ \text{s.t.} \quad & \dot{x}(t) = (1 - x(t) - \rho)(x(t) + p_c^*(\alpha(t))\rho)(1 - \alpha(t))q - \alpha(t)x(t)(1 - x(t) + (1 - p_c^*(\alpha(t)))\rho)p.\end{aligned} \tag{3.18}$$

Writing the Hamiltonian for this problem,

$$\begin{aligned} H(x, \alpha, \lambda) &= g(x, \alpha) + \lambda f(x, \alpha) \\ &= (c_1 - c_2 - c_3)\alpha(x + p_c^*(\alpha)\rho) + c_2\alpha + c_3(x + p_c^*(\alpha)\rho) \\ &+ \lambda\Big((1 - x - \rho)(x + p_c^*(\alpha)\rho)(1 - \alpha)q - \alpha x(1 - x + (1 - p_c^*(\alpha))\rho)p\Big). \end{aligned} \tag{3.19}$$

This is a more difficult problem than the one in (3.6). 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 3.8.1. 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 3.8.2 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 3.8.3 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 3.8.2. *Let $\alpha^*(t)$ denote the principal's optimal audit rate. If $\rho < x_N$, then $\alpha(t)^* < \alpha_N$.*

Proof. See Appendix. □

Proposition 3.8.3. *If $\rho > x_N$, then $\alpha(t)^* \geq \alpha_N$.*

Proof. See Appendix. □

3.9 Discussion

This chapter presents 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 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 chapter 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 chapter. For example, 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 chapter 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 chapter. 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.

Chapter 4

Best Response and Fictitious Play Learning

4.1 Introduction

The previous chapter focused on the social learning aspects of decision making through employing a model based on replicator dynamics. This model endowed the agents with no strategic abilities, instead, agents mechanically copied strategies that seemed more successful. This chapter considers another kind of learning that departs from the replicator idea and equips the agents with some limited ability for strategic reasoning. I also consider a discrete time model as opposed to the continuous time framework of the previous chapter. The main setting is still the same: agents play the Cheat-Audit game against a principal, but learn in a different manner.

The learning model I consider in this chapter is based on Fictitious Play (Brown (1951)), where agents adapt their strategies based on all or a truncated portion of history and how the game was played in the past. Each period, agents observe a function of history and choose their actions accordingly, after which they exit the system and are replaced by another batch of agents in the next round. The assumption that agents are replaced each round allows us to avoid worrying about repeated interaction concerns. I address a similar question to the one asked in the previous chapter, namely, how should the principal play the game given the learning dynamics? I provide analytical and computational answers to this question. I first examine a version of the problem where agents adjust their actions based on play in the previous round only, and provide the optimal policies for

this case. I then show computationally that policies with a similar structure are also optimal —in a computational sense— for the case when agents look further back into the past.

Like prior results, some of the work presented here has direct implications on the understanding of conventions: if a game has been played in a certain way for a long period of time, is it possible for a principal to change how newer generations play it? The main difference between the inquiries presented here and results like Young (1993) is that I do not focus on whether a game converges to an equilibrium, but rather on whether it is possible to direct play in a way that is more beneficial to a principal, who can either be a selfish or a benevolent (social) planner.

4.2 Model

Consider a long-lived principal that repeatedly plays the Cheat-Audit game against short-lived agents. In each period, a large number of agents play the game for exactly one period and then exit the system forever. Agents therefore are only interested in maximizing their own payoffs during this one period, and they choose their actions based on some function of how the game was played in the past. I will start with a limiting case of fictitious play, where agents best respond to the immediate past, taking only the last period of play into account. In round t, the principal decides the audit rate α_t and the game is played. The principal then adjusts α_{t+1} based on how the agents respond and the strategy that he is implementing.

4.2.1 Learning Dynamics

A strategy profile at time t is denoted by σ_t, where $\sigma_t = (x_t, \alpha_t)$, x_t and α_t are the fraction of C players and the audit rate in period t, respectively. History at the beginning of period t is denoted by $h_t = g(\sigma_{t-k+1}, ..., \sigma_{t-1})$ and is a function of how the game was played in the past k out of $t-1$ periods, with the possibility that $k = t-1$ so that all past periods are taken into account. Players use the information in h_t to decide on their choice of action in period t, giving rise to the action profile $\sigma_t = (x_t, \alpha_t)$. Because I will only consider learning on the side of the agents and not the principal, I will use h_t to refer to how the principal played the game in the past, so that $h_t = g(\alpha_1, ..., \alpha_{t-1})$.

Agents utilize a rule $a(h_t)$ that prescribes how they should play when faced with a certain history. Similarly, the principal's strategy is given by $b(h_t)$. If one thinks about $a(h_t)$ as the probability with which an agent plays action C when facing history h_t, then $\sigma_t = (x_t, \alpha_t) = (a(h_t), b(h_t))$ describes the aggregate action of the population and the action of the principal, respectively, at time t. The *successor* to state h_t, h_{t+1}, is produced at the end of period t via a transition function $f(\sigma_t, h_t)$. The function f depends on the learning model under consideration. The following are examples of h_t for the two learning models that I consider in this chapter.

Best response: In best response, agents look at the action played in the last period and best respond to it. Here, h_t is simply equal to α_t.

Fictitious play: In fictitious play, each player keeps a running average of how his opponent played the game in all past periods. This information is recorded in h_t, so that $h_t = g(\alpha_1, ..., \alpha_{t-1}) = \frac{\sum_{i=1}^{t-1} \alpha_i}{t-1}$, where the function g simply averages its arguments. The players then best respond in period $t+1$ to this running average.

4.3 Analysis

Before starting the analysis, I present a few useful and straightforward results that will help in comparing the performance of different strategies later on in the chapter. The following is a simple calculation of the cost of repeatedly playing the Cheat-Audit game in a fully rational setting.

Proposition 4.3.1. *The cost to the principal from playing the Cheat-Audit game in a fully rational setting is equal to* $\frac{1}{1-\delta}\frac{c_2 c_3}{c_3+c_2-c_1}$*, where* $0 < \delta < 1$ *is a discount factor.*

Proof. The unique Nash equilibrium strategy profile, σ_N, is given by $\sigma_N = (x_N, \alpha_N) = (\frac{c_2}{c_3+c_2-c_1}, \frac{v_3-v_2}{v_1})$. The per-stage cost is given by $(c_1 - c_2 + c_3)\alpha x + c_2\alpha + c_3 x$, which translates to $\frac{c_2 c_3}{c_3+c_2-c_1}$ at σ_N. Assuming a discount factor of δ and summing this quantity over the entire horizon, one gets $\sum_{t=0}^{\infty} \delta^t \frac{c_2 c_3}{c_3+c_2-c_1} = \frac{1}{1-\delta}\frac{c_2 c_3}{c_3+c_2-c_1}$. □

Because each agent plays the game only once, they are necessarily strategically myopic, optimizing only for the period they play in without taking future play into account. Additionally, the

learning models considered here do not allow agents to speculate on the strategic abilities of the principal; agents simply assume that the principal audits with a certain rate that they try to approximate via history, and then best respond to. Let $a(h_t)$ denote the probability with which an agent plays action C as a function of h_t, then we have the following simple observation about the learning rule for fictitious play

Observation 4.3.2. *In the beginning of period t, an agent facing history h_t plays action C with probability $a(h_t)$, where*

$$a(h_t) = \begin{cases} 1, & h_t < \alpha_N; \\ 0, & h_t > \alpha_N; \\ y, & h_t = \alpha_N. \end{cases} \tag{4.1}$$

Simply stated, if an agent believes that the audit rate is low enough such that cheating is profitable when compared to the risk of being audited, then he will play C. The opposite is true when the audit rate is thought to be high enough to not make playing C an attractive option. If the approximation of the audit rate is exactly equal to the Nash audit rate then the agent is indifferent between the two actions, in which case I assume he just plays C with some probability y.

In the following I will examine two strategies. The first is called a threshold strategy, which only audits when history hits a certain threshold. The second is called the over-audit strategy: Play $\alpha_N + \epsilon$ in every period, so that $x_t = 0$ for all t. This strategy has a cost-per-stage of $c_2(\alpha_N + \epsilon)$, for a total cost of $\frac{1}{1-\delta} c_2(\alpha_N + \epsilon)$. The reason for choosing these two strategies is the following intuition. In the threshold strategy, as long as the running history is less than the Nash equilibrium audit rate, then the principal can obtain his most preferred payoff by not auditing and also having the population not cheating. When history finally crosses a threshold, the principal audits and the cycle starts again. One possibility that would make this strategy not optimal is if the Nash audit rate or the cost c_2 of auditing an honest agent is very low, so that the principal can keep the cheating population at 0 forever by always auditing slightly above the Nash and not incurring a high enough cost from auditing honest agents.

4.3.1 Best Response

I consider best response dynamics as a special case of fictitious play when agents do not take into account the whole history but instead only use the last period of play as a guide to how to play in the current period according to (4.1). This way, the principal's decision in period t affects his payoffs in t and $t+1$ only. I will analyze this case by writing the dynamic program for the infinite horizon version of the problem. The optimal policy will be derived through a sequence of lemmas, and the following definitions and assumptions will prove useful in the derivations of this section.

Definition 4.3.3. *A* threshold strategy α^{th} *is given by*

$$\alpha^{th} = \begin{cases} 1, & h_t < \alpha_N \text{ or } h_t = \alpha_N \text{ and } y > x_N; \\ 0, & h_t > \alpha_N \text{ or } h_t = \alpha_N \text{ and } y < x_N. \end{cases} \tag{4.2}$$

The abundance of parameters in the problem and the way these parameters interact with each other makes it difficult to identify a policy that is optimal with no restrictions on the values of these parameters or how they relate to one other. Because of this, the analysis I provide is divided into several sections, addressing the possible scenarios that arise in the different regions of the parameter space. For a large part of this space, I will show that the strategy in Definition 4.2 is optimal. Sometimes no optimal strategy exists for the principal. I will isolate these cases and show when they arise based on the values of the problem. Let us start with an assumption on the discount factor δ.

Assumption 4.3.4. *The discount factor* δ *satisfies* $\delta < \frac{c_2\alpha_N}{c_1 - c_2\alpha_N}$, *where* α_N *is the Nash audit rate.*

The Nash audit rate α_N has a strong effect on how the game should be played. The following assumption will be useful in dissecting that effect. The optimal policy is significantly different when the assumption is satisfied than when it is not.

Assumption 4.3.5. *The Nash audit rate* α_N *satisfies* $\alpha_N > \frac{c_1}{2c_2}$.

I will use the standard infinite horizon dynamic programming terminology in analyzing the game. Since in the best response case history in period $t+1$ is simply the action that was taken by the

principal in period t, I can use Observation 4.3.2 to slightly abuse notation and think about the state as x_t, since $h_t = \alpha_{t-1}$ and α_{t-1} maps to exactly one value of x_t. Dropping the time subscript, a state x can have only one of three possible values, 0, 1, or y, and the transition function $f(h_t, \sigma_t)$ simplifies to $f(\alpha)$ where

$$f(\alpha) = \begin{cases} 1, & \alpha < \alpha_N; \\ 0, & \alpha > \alpha_N; \\ y, & \alpha = \alpha_N. \end{cases} \tag{4.3}$$

We can now write the optimal value function $v(x)$ as $v(x) = u(x, \alpha) + \delta v(f(\alpha))$, where $u(x, \alpha)$ is the one-period cost when the state is x and the action is α and $0 \leq \delta < 1$ is a discount factor. Let $v_{th}(x)$ indicate the payoff from following the threshold strategy α_{th} when the state is x, so that $v_{th}(x) = u_{th}(x) + v_{th}(f(\alpha_{th}))$. Under the assumptions above and using the one-step deviation principle, I will show that the strategy α_{th} is optimal. Similar to the previous chapter, the one-period cost function is given by $u(x, \alpha) = (c_1 - c_2 - c_3)\alpha x + c_2\alpha + c_3 x$. Therefore, using the strategy in Definition 4.3.3 to determine the value of α, we can write $u_{th}(x)$ as

$$u_{th}(x) = \begin{cases} 0, & x = 0; \\ c_1, & x = 1; \\ c_3 y, & x = y \text{ and } y < x_N; \\ (c_1 - c_2)y + c_2, & x = y \text{ and } y > x_N. \end{cases} \tag{4.4}$$

The next lemma orders $v_{th}(x)$ based on the values of x.

Lemma 4.3.6. *Let $y > x_N$, then $v_{th}(0) < v_{th}(1) < v_{th}(y)$.*

Proof. Using Equation (4.2), we can write $v_{th}(0)$ as

$$v_{th}(0) = u_{th}(0) + \delta v_{th}(1)$$

and

$$v_{th}(1) = u_{th}(1) + \delta v_{th}(0)$$

and, by using (4.4)

$$v_{th}(0) = \frac{u_{th}(0) + \delta u_{th}(1)}{1-\delta^2} = \frac{\delta u_{th}(1)}{1-\delta^2} = \frac{\delta c_1}{1-\delta^2} \tag{4.5}$$

and

$$v_{th}(1) = \frac{u_{th}(1) + \delta u_{th}(0)}{1-\delta^2} = \frac{u_{th}(1)}{1-\delta^2} = \frac{c_1}{1-\delta^2}. \tag{4.6}$$

Since $0 < \delta < 1$, it follows that $v_{th}(0) < v_{th}(1)$.

Now write $v_{th}(y)$

$$v_{th}(y) = u_{th}(y) + \delta v_{th}(0) = (c_1 - c_2)y + c_2 + \delta v_{th}(0). \tag{4.7}$$

Consider the inequality

$$\frac{\delta}{1+\delta} < \frac{(c_1 - c_2)y + c_2}{c_1}$$

which is always true since the RHS is bounded below by 1 as $0 < c_1 < c_2$ and $y \in [0, 1]$ and the LHS is bounded above by 1. Manipulating this inequality, we have

$$\frac{\delta c_1}{1+\delta} < (c_1 - c_2)y + c_2$$

. Multiplying the LHS by $\frac{1-\delta}{1-\delta}$, using $1 - \delta^2 = (1-\delta)(1+\delta)$ and rearranging:

$$\frac{\delta c_1}{1-\delta^2} < \frac{(c_1 - c_2)y + c_2}{1-\delta}.$$

From (4.5), the LHS is equal to $v_{th}(0)$ and

$$v_{th}(0)(1-\delta) < (c_1 - c_2)y + c_2.$$

Finally, rearranging the above, we get

$$v_{th}(0) < (c_1 - c_2)y + c_2 + \delta v_{th}(0).$$

The RHS of the above is the same as (4.7), giving

$$v_{th}(0) < v_{th}(y).$$

Noting that the difference between $v_{th}(0)$ and $v_{th}(1)$ is a factor of δ, we start from the inequality $\frac{\delta^2}{1+\delta} < \frac{(c_1-c_2)y+c_2}{c_1}$, which is always true, as the LHS is bounded above by $\frac{1}{2}$, and progress in the same manner as above to yield $v_{th}(1) < v_{th}(y)$. □

The next step is to rank the same quantities when $y < x_N$.

Lemma 4.3.7. *Let $y < x_N$, then $v_{th}(0) < v_{th}(1)$ and $v_{th}(0) < v_{th}(y)$.*

Proof. From the proof of Lemma 4.3.6, $v_{th}(0) < v_{th}(1)$ regardless of the value of y. Writing $v_{th}(y)$ for the case when $y < x_N$, the threshold strategy sets $\alpha_{th} = 0$ and we have

$$\begin{aligned} v_{th}(y) = u(y) + \delta v_{th}(1) &= c_3 y + \delta \frac{c_1}{1-\delta^2} \\ &= c_3 y + \delta v_{th}(0) \\ &> v_{th}(0). \end{aligned} \tag{4.8}$$

□

Proposition 4.3.8. *Let one or both of Assumptions 4.3.4 and 4.3.5 hold, then the threshold strategy is the optimal strategy for the principal when agents are playing according to best response.*

Proof. I will prove the result by showing that one-step deviations from the threshold strategy under the assumption(s) in the statement of the proposition do not lead to reductions in the cost. For this to be the case we have to examine deviations in the three possible states, $x = 1$, $x = 0$, and $x = y$. Notice that the threshold strategy never leads to state y, so checking deviations in that state is only to cover the scenario when the system starts in state y.

Case 1: $x = 1$: The principal's problem when $x = 1$ is

$$\min_{\alpha} u(1, \alpha) + \delta v_{th}(f(\alpha))$$
$$= \min_{\alpha} (c_1 - c_2 - c_3)\alpha + c_2\alpha + c_3 + \delta v_{th}(f(\alpha)). \tag{4.9}$$

The threshold strategy sets $\alpha_{th} = 1$ in this case, giving

$$c_1 + \delta v_{th}(0). \tag{4.10}$$

Consider setting $\alpha = \bar{\alpha}$ where $\alpha_N < \bar{\alpha} < 1$. By (4.3), the transition resulting from this setting is to state 0, then (4.9) becomes

$$(c_1 - c_2 - c_3)\bar{\alpha} + c_2\bar{\alpha} + c_3 + \delta v_{th}(0).$$

The first part of this expression is a linear function of α that is minimized by setting $\alpha = 1$, as in (4.10). The second part, $\delta v_{th}(0)$ is the same in (4.10) and in the expression above, hence any deviation that sets $\alpha \in (\alpha_N, 1)$ cannot improve the cost.

Now consider setting $\alpha = \underline{\alpha}$ where $0 \leq \underline{\alpha} < \alpha_N$. Using (4.3), Equation (4.9) becomes

$$(c_1 - c_2 - c_3)\underline{\alpha} + c_2\underline{\alpha} + c_3 + \delta v_{th}(1).$$

The same reasoning as above applies: the first part is a linear function in α that is minimized as in (4.10). The second part in the above expression is $\delta v_{th}(1)$, which by Lemmas 4.3.6 and 4.3.7 has higher cost when compared to $\delta v_{th}(0)$, and therefore a deviation that sets $\alpha \in [0, \alpha_N)$ is not cost-reducing.

Finally, consider the case when setting $\alpha = \alpha_N$ in (4.9), then we have

$$(c_1 - c_2 - c_3)\alpha_N + c_2\alpha_N + c_3 + \delta v_{th}(y).$$

Under Assumption 4.3.5, Lemmas 4.3.6 and 4.3.7 apply and hence $v_{th}(y) > v(0)$. Using the same reasoning as above it is straightforward to show that the expression above has higher cost than (4.10). This concludes the analysis for the $x = 1$ case.

Case 2: $x = 0$: The principal's problem is

$$\begin{aligned} &\min_{\alpha} u(0,\alpha) + \delta v_{th}(f(\alpha)) \\ &= \min_{\alpha} c_2\alpha + \delta v_{th}(f(\alpha)). \end{aligned} \tag{4.11}$$

The threshold strategy sets $\alpha_{th} = 0$ in this case, giving

$$\delta v_{th}(1). \tag{4.12}$$

Proceeding as before, consider a strategy that sets $\alpha = \underline{\alpha}$ where $0 < \underline{\alpha} < \alpha_N$. From (4.3), this strategy has cost equal to

$$c_2\underline{\alpha} + \delta v_{th}(1) \tag{4.13}$$

which is clearly higher than (4.12). Now let $\alpha = \bar{\alpha}$ where $\bar{\alpha} > \alpha_N$, then (4.11) becomes

$$c_2\bar{\alpha} + \delta v_{th}(0). \tag{4.14}$$

This is bigger than (4.12) when

$$\bar{\alpha} > \frac{\delta(v_{th}(1) - v_{th}(0))}{c_2} = \frac{\delta c_1(1-\delta)}{c_2(1-\delta^2)} = \frac{\delta c_1}{c_2(1+\delta)}.$$

Taking this inequality with the fact that $\alpha > \alpha_N$, we get that the threshold policy is still optimal when

$$\frac{\delta c_1}{c_2(1+\delta)} < \alpha_N \tag{4.15}$$

since if that was not the case then an improvement in cost can be obtained by setting α to some value

in $(\alpha_N, \bar{\alpha})$. For the inequality in (4.15) to hold with the constraints that $0 < c_1 < c_2$, $0 \leq \alpha_N \leq 1$, and $0 < \delta < 1$, one of the following conditions must be met: Either $\alpha_N > \frac{c_1}{2c_2}$, or $\delta < \frac{\alpha_N c_2}{c_1 - \alpha_N c_2}$, which is true under the assumptions in the statement of the proposition.

Finally consider the case $\alpha = \alpha_N$, in which case the cost becomes

$$c_2\alpha_N + \delta v_{th}(y). \tag{4.16}$$

Compare this quantity with (4.14). There is an $\bar{\alpha} > \alpha_N$ for which (4.14) is always smaller than (4.16). Because by Lemmas 4.3.6 and 4.3.7, $v_{th}(0) < v_{th}(y)$, we can choose $\bar{\alpha} = \alpha_N + \epsilon$ where $0 < \epsilon < \frac{v_{th}(y) - v_{th}(0)}{c_2}$, i.e., setting α in the region $\alpha > \alpha_N$ dominates setting $\alpha = \alpha_N$. But we have already shown that the threshold strategy dominates setting $\alpha > \alpha_N$, and hence the cost of (4.16) cannot improve on (4.12) and the threshold strategy is still optimal.

To finish the proof, we turn to the remaining possible value of x.

Case 3: $x = y$: Assume $y > x_N$, the principal's problem is

$$\begin{aligned} &\min_{\alpha} u(y, \alpha) + \delta v_{th}(f(\alpha)) \\ &= \min_{\alpha} (c_1 - c_2 - c_3)\alpha y + c_2\alpha + c_3 y + \delta v_{th}(f(\alpha)). \end{aligned} \tag{4.17}$$

The threshold strategy sets $\alpha_{th} = 1$ in this case, giving

$$c_2 + (c_1 - c_2)y + \delta v_{th}(0). \tag{4.18}$$

Consider setting $\alpha = \bar{\alpha}$ where $\bar{\alpha} > \alpha_N$. This gives

$$(c_1 - c_2 - c_3)\bar{\alpha} y + c_2\bar{\alpha} + c_3 y + \delta v_{th}(0).$$

Since the first part of (4.18) is optimal solution to $u(y, \alpha)$ and the second part of the above expression, $v_{th}(0)$, is the same as (4.18), $\bar{\alpha}$ is not cost-improving.

Next, turn to the case where $\alpha = \underline{\alpha}$ where $0 < \underline{\alpha} < \alpha_N$. This is obviously not cost improving since (4.17) becomes

$$(c_1 - c_2 - c_3)\underline{\alpha}y + c_2\underline{\alpha} + c_3 y + \delta v_{th}(1)$$

and both the first component has higher cost than (4.18) and the second has $v_{th}(1) > v_{th}(0)$.

When $\alpha = \alpha_N$, the case is similar to setting $\alpha > \alpha_N$ but with $v_{th}(1)$ replaced by $v_{th}(y)$ to give a total cost of

$$(c_1 - c_2 - c_3)\alpha_N y + c_2\alpha_N + c_3 y + \delta v_{th}(y),$$

which is again higher than (4.18).

When $y < x_N$, the cost for the threshold strategy, which sets $\alpha = 0$ is

$$c_3 y + \delta v_{th}(1). \tag{4.19}$$

Any strategy that sets $0 < \underline{\alpha} < \alpha_N$ cannot improve on the cost of (4.19). The cost for $\underline{\alpha}$ is given by

$$(c_1 - c_2 - c_3)\underline{\alpha}y + c_2\underline{\alpha} + c_3 y + \delta v_{th}(1)$$

since the linear part of the principal's problem is minimized by setting $\alpha = 0$ and the second part in (4.19) is the same as in the expression directly above.

Consider $\bar{\alpha} > \alpha_N$, then the cost is

$$(c_1 - c_2 - c_3)\bar{\alpha}y + c_2\bar{\alpha} + c_3 y + \delta v_{th}(0)$$

which is bigger than (4.19) when

$$\bar{\alpha} > \frac{\delta(v_{th}(1) - v_{th}(0))}{(c_1 - c_2 - c_3)y + c_2} = \frac{\delta c_1(1-\delta)}{((c_1 - c_2 - c_3)y + c_2)(1-\delta^2)} = \frac{\delta c_1}{((c_1 - c_2 - c_3)y + c_2)(1+\delta)}$$

which approaches (4.15) as $y \to 0$. The case when $\alpha = \alpha_N$ is treated in the same manner as in the

proof of the similar case when $x = 0$. Because it is always possible to find an ϵ such that $\alpha = \alpha + \epsilon$ dominates $\alpha = \alpha_N$, and because $\alpha = \alpha + \epsilon$ itself is dominated by $\alpha = 0$ in this scenario, $\alpha = \alpha_N$ is dominated by $\alpha = 0$. □

What happens when Assumption 4.3.4 is violated? To answer this question I first give the following definition

Definition 4.3.9. *An* over-audit strategy α^{oa} *is given by*

$$\alpha^{oa} = \begin{cases} 1, & h_t < \alpha_N \text{ or } h_t = \alpha_N \text{ and } y > x_N; \\ \alpha_N + \epsilon, & h_t > \alpha_N \text{ or } h_t = \alpha_N \text{ and } y < x_N. \end{cases} \tag{4.20}$$

Note that as soon as $h_t > \alpha_N$, α^{oa} is equal to $c_2 + \epsilon$ forever. Consider now the proof of optimality for the threshold strategy, for the case when $x = 1$ no assumptions on any of the parameters were needed: it is always optimal to set $\alpha = 1$ when $x = 1$. For the cases when $x = 0$ we had to use the assumption on δ and/or α_N because otherwise choosing α equal to any value in the interval between α_N and $\frac{\delta c_1}{c_2(1+\delta)}$ reduces the cost from the threshold strategy. It is exactly under these conditions where the over-audit strategy dominates the threshold strategy. However, the over-audit strategy is not optimal. In fact, no optimal strategy exists in this case, since one can always reduce the cost by choosing a smaller ϵ. As $\epsilon \to 0$, the cost approaches $\frac{c_2\alpha_N}{1-\delta}$.

The following result formalizes the preceding discussion.

Proposition 4.3.10. *When neither Assumption 4.3.4 or 4.3.5 hold, the over-audit strategy is undominated by any other strategy.*

The proof of the proposition is similar to the proof of Proposition 4.3.8. Let $v_{oa}(x)$ denote the payoff from following the over-audit strategy when the state is x. Before beginning the proof, I rank $v_{oa}(0)$ and $v_{oa}(1)$.

Lemma 4.3.11. *For the over-audit strategy,* $v_{oa}(0) < v_{oa}(1)$.

Proof. First write $v_{oa}(0)$

$$v_{oa}(0) = c_2\alpha^{oa} + \delta v_{oa}(0) \tag{4.21}$$

where $\alpha^{oa} = \alpha_N + xe$. Write $v_{oa}(1)$ as

$$v_{oa}(1) = c_1 + \delta v_{oa}(0). \tag{4.22}$$

From (4.21), $v_{oa}(0) = \frac{c_2\alpha^{oa}}{1-\delta}$, which is what I have found before by summing the cost over the entire horizon. I can now rewrite (4.22) as

$$v_{oa}(1) = c_1 + \delta\frac{c_2\alpha^{oa}}{1-\delta}.$$

This is bigger than (4.21) when

$$c_1 + \delta\frac{c_2\alpha^{oa}}{1-\delta} > \frac{c_2\alpha^{oa}}{1-\delta}$$

which happens when $\alpha^{oa} < \frac{c_1}{c_2}$. Since $\frac{c_1}{c_2} > \frac{c_1}{2c_2}$ and $\alpha_N < \frac{c_1}{2c_2}$, we have $\alpha_N < \frac{c_1}{c_2}$. Since one can always find $\epsilon > 0$ such that $\alpha^{oa} = \alpha_N + \epsilon < \frac{c_1}{c_2}$, the lemma is proved. □

Now I prove Proposition 4.3.10.

Proof. Using the one-step deviation principle as before:

Case 1: $x = 1$: $v_{oa}(1)$ is given by (4.22). Consider $0 \leq \underline{\alpha} < \alpha_N$, this gives

$$(c_1 - c_2 - c_3)\underline{\alpha} + c_2\underline{\alpha} + c_3 + \delta v_{oa}(1)$$

which is higher than (4.22) since the linear part is minimized at $\alpha = 1$ and $v_{oa}(1) > v_{oa}(0)$. Consider $\bar{\alpha} > \alpha_N$, then the cost becomes

$$(c_1 - c_2 - c_3)\bar{\alpha} + c_2\bar{\alpha} + c_3 + \delta v_{oa}(1)$$

which is minimized at $\alpha = 1$, hence $v_{oa}(1)$ is undominated.

Case 2: $x = 0$: $v_{oa}(0)$ is given by (4.21) and is equal to $\frac{c_2\alpha^{oa}}{1-\delta}$. Consider setting $0 \leq \underline{\alpha} < \alpha_N$;

this gives

$$c_2\underline{\alpha} + \delta v_{oa}(1).$$

Since the linear part of this equation is minimized at $\underline{\alpha} = 0$, I will focus on this case, since any $\alpha < \alpha_N$ will have higher cost. When $\underline{\alpha} = 0$, the cost is simply $\delta v_{oa}(1)$. This is higher than $v_{oa}(0)$ when

$$\delta > \frac{v_{oa}(0)}{v_{oa}(1)} = \frac{v_{oa}(0)}{c_1 + \delta v_{oa}(0)}. \tag{4.23}$$

Replacing $v_{oa}(0)$ with $\frac{c_2\alpha^{oa}}{1-\delta}$ and solving the inequality for δ after noting that $0 < c_1 < c_2$ and $0 \leq \alpha^{oa} \leq 1$, I find that (4.23) is satisfied when $0 < \alpha^{oa} < \frac{c_1}{2c_2}$ and $\delta > \frac{c_2\alpha^{oa}}{c_1 - c_2\alpha^{oa}}$, which are the conditions ruled out by Assumptions 4.3.4 and 4.3.5. Since the statement of the proposition assumes that neither holds, the conditions for (4.23) are satisfied.

Finally, what happens when we set α to $\bar{\alpha} > \alpha_N$ in state $x = 0$? The cost is given by

$$c_2\bar{\alpha} + \delta v_{oa}(0).$$

This is only higher than (4.21) when $\bar{\alpha} > \alpha^{oa}$, but one can set $\bar{\alpha}$ to be less than α^{oa} while still being higher than α_N. This means that, while the over-audit strategy is not dominated by another strategy, it also does not provide an optimal solution, since the principal can always reduce the amount of auditing while still being above the α_N threshold.

□

4.4 Simulation

This section carries forward the ideas presented so far through simulation and numerical experiments. I will first check the theoretical results obtained above with a simulation of the Cheat-Audit game when agents are playing according to best response. The code for all simulations can be found in the appendix. To simplify the discussion, I assume that the first period, where there is no history yet, starts with no cheating activity and full auditing activity. We can of course think about and simulate

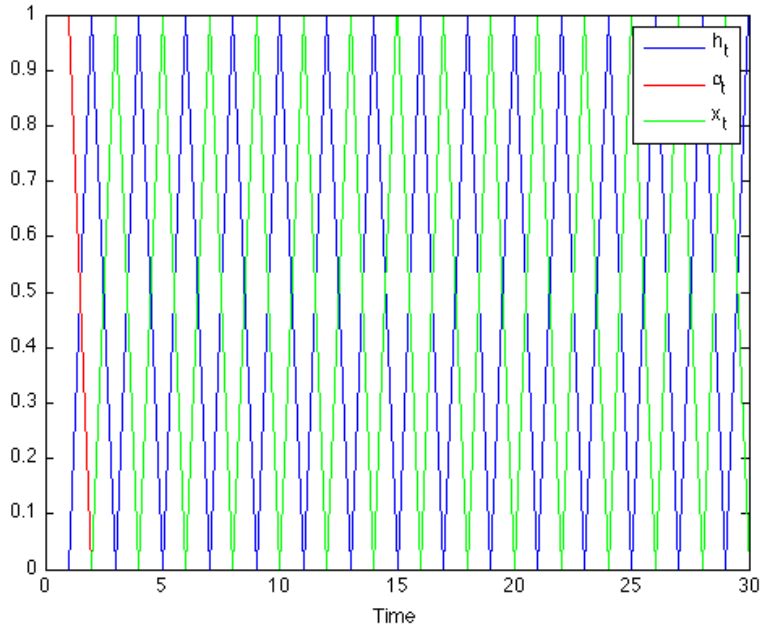


Figure 4.1: Best Response with $c_1 = 1, c_2 = 3, c_3 = 10$, and $\alpha_N = 0.25$

scenarios other than this one. For example, assume that agents start by playing y when no history is available and that $y > x_N$. The optimal solution according to the simulations then still takes the same form (for example, cycles) and retains the same long-run properties (for example, convergence of the number of periods where auditing happens to the Nash audit rate in fictitious play), but the periods on which certain values, like $\alpha = 1$ and $x = 1$, differ depend on the assumptions about the first period.

Consider the following set of parameters: $c_1 = 1$, $c_2 = 3$, $c_3 = 10$, and $\alpha_N = 0.25$ (the costs imply that the Nash rate for cheating is also equal to 0.25). Proposition 4.3.8 tells us that no matter what δ is, the optimal policy for these parameters is the threshold strategy. The reason is that α_N satisfies Assumption 4.3.5 and hence Proposition 4.3.8 applies. Figure 4.1 shows the optimal policy for a 30-period run. Note of course that the simulation is an approximation of the optimal policy, since it runs for a finite horizon (a fixed number of rounds). The auditing action, given by the red lines, coincides with the cheating activity, given by the green lines (because of this the red lines do not appear in the figure). h_t is given by the blue lines. Notice how both x_t and α_t are set to 1 when

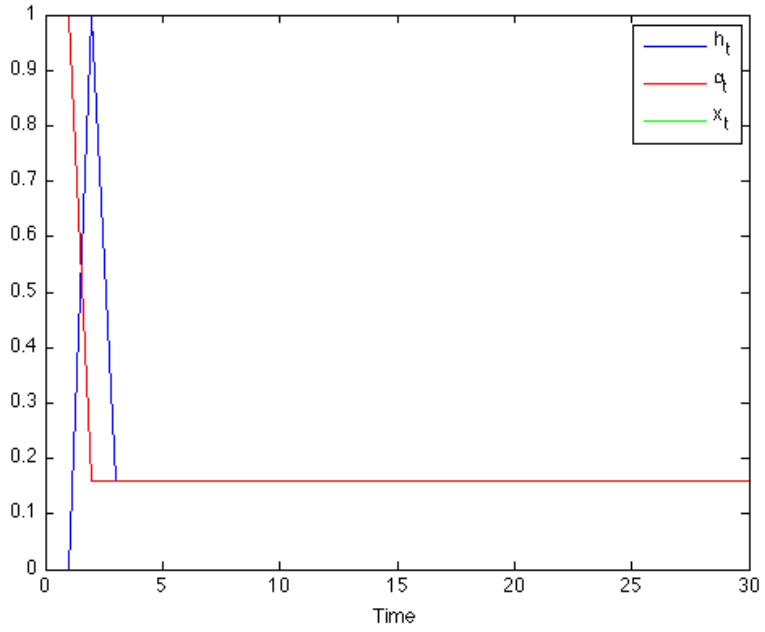


Figure 4.2: Best response with $c_1 = 1, c_2 = 3, c_3 = 10$, $\delta = 0.93$, and $\alpha_N = 0.15$

$h_t = 0$ and vice versa, as the theory implies.

I now satisfy the condition on α_N by setting it equal to a value in $[0, \frac{c_1}{2c_2}) = [0, 1/6)$. Suppose $\alpha_N = 0.15$, then $\delta < \frac{c_2\alpha_N}{c_1 - c_2\alpha_N} = 0.8182$ implies that the threshold strategy is still optimal, as indeed it is. Setting δ above the threshold, the simulation indicates that the over-audit strategy is the optimal solution, as seen in Figure 4.2. Since the code searches for an optimal α in increments of 0.01, the optimal α selected is the first increment after the value of α_N, in this case $\alpha = 0.16$, just like the over-audit strategy predicts. Of course, this is the optimal solution when α is restricted to take values in this discrete domain, but in theory we can still reduce α further —but still above $\alpha_N = 0.15$— and obtain less cost. In the figure, $x_t = 0$ for all t, and hence does not appear in the figure.

How do the results change if agents, instead of learning only from the most recent period, take into account all of history, as in the standard fictitious play model? The transition function $f(\sigma_t, h_t)$ in this case is given by $f(\alpha_t, h_t) = \frac{th_t + \alpha_t}{t+1}$, which indicates that actions taken by the principal late in the game have relatively little impact on history and subsequently on how agents will play. One

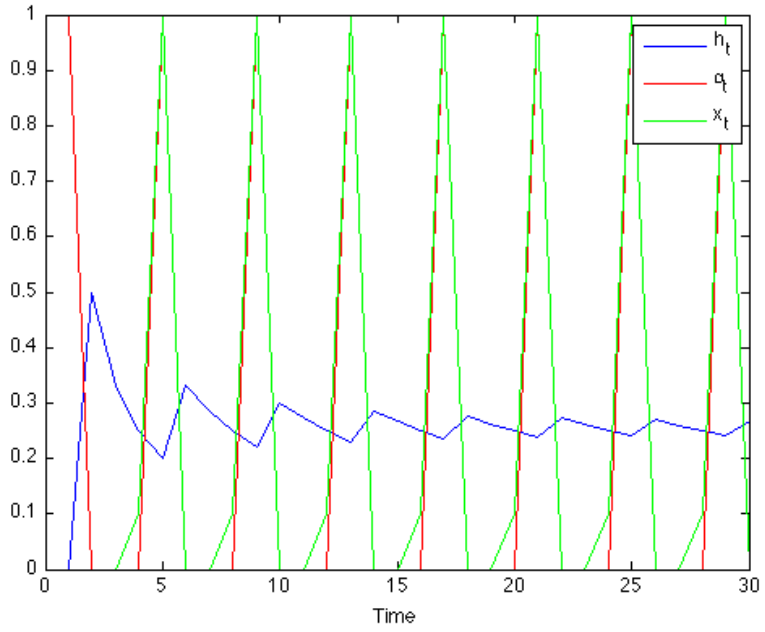


Figure 4.3: Fictitious play with $c_1 = 1, c_2 = 3, c_3 = 10$, $y = 0.1$, $\delta = 0.6$, and $\alpha_N = 0.25$

can think of this as the agents sticking to a certain convention if the game has been played for a long period of time. This means that a principal who decides at some point farther down the game that it would like to influence or change how the agents play has a difficult road ahead.

Performing the same experiment with the parameters from Figure 4.1 but with agents taking all past history into account, we get Figure 4.3.

Again, the green and red lines coincide, which leads to the first observation about this figure, which is that the cheating pattern follows the auditing pattern, so that there is a synchronization between cheating and auditing: periods where no auditing is happening are also periods where no or little cheating is taking place. When an audit happens, it is in a period where there is enough cheating to render the audit useful. The second observation is that the auditing behavior is cyclical and the period of the cycle is equal to the Nash audit rate. Because the problem takes place in a discrete time setting, the period of the cycle approximates α_N, so that the fraction of periods with $\alpha = 1$ approaches α_N as $t \to \infty$. This is the outcome that one would get when following the strategy in Definition 4.2, since if this strategy is followed from the beginning of the game, then each period

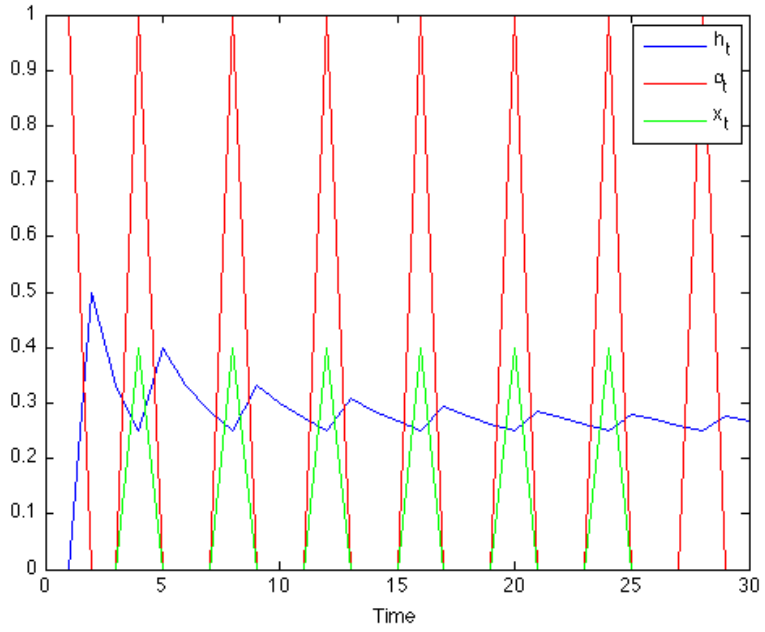


Figure 4.4: Fictitious Play with $c_1 = 1, c_2 = 3, c_3 = 10$, $y = 0.4$, $\delta = 0.6$, and $\alpha_N = 0.25$

with $\alpha = 1$ sustains an average of $\frac{1}{\alpha_N} - 1$ periods with $\alpha = 0$ before h_t returns to a value that is higher than α_N, at which point the strategy sets $\alpha = 1$ again and the pattern is repeated. As $t \to \infty, h_t \to \alpha_N$. As can be seen in the figure, even for a run of only 30 periods, we can see h_t starting to converge to the Nash audit rate.

One main difference between the threshold strategy under fictitious play and best response is that under fictitious play there is a chance that $h_t = \alpha_N$, where in best response h_t was always either 0 or 1. Consequently, the value of y does not matter on the path of the optimal solution in best response, since the system is never in state y as long as $0 < \alpha_N < 1$. In contrast, fictitious play has periods where the system is in state y and, depending on the value of y, the principal acts differently. Figure 4.3 and Figure 4.4 illustrate this point. All the parameters are set exactly the same except in the first figure $y = 0.1 < x_N$ and in the second $y = 0.4 > x_N$. If we look at the curve of cheating activity in the first figure, there is a wedge whenever the period number is a multiple of 4. This is because on these periods, and given the value of $\alpha_N = 0.25$, h_t is exactly equal to α_N. Because the value of y is low, the principal can afford to let some cheating happen without

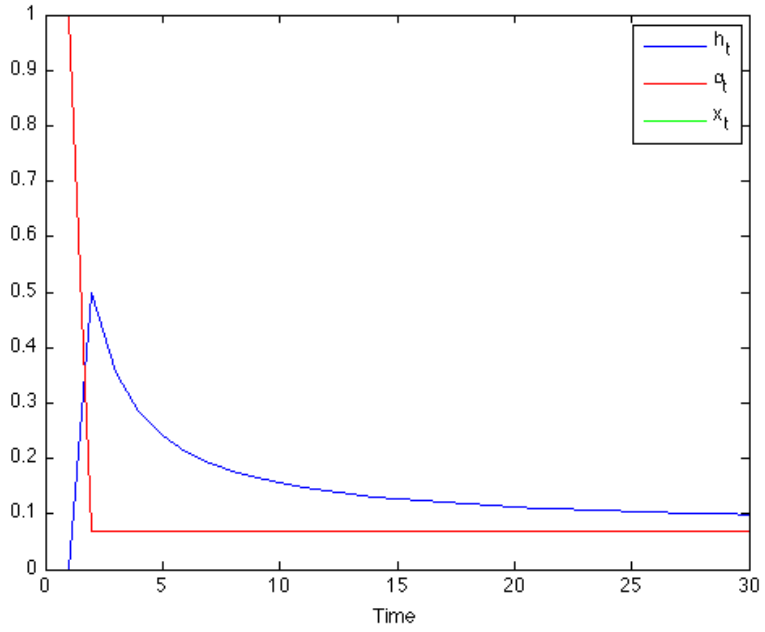


Figure 4.5: Fictitious Play with $c_1 = 1, c_2 = 3, c_3 = 10$, $y = 0.4$, $\delta = 0.6$, and $\alpha_N = 0.068$

auditing. But since having $h_t = \alpha_N$ and $\alpha_t = 0$ means that in period $t+1$ $h_t < \alpha_N$, both cheating and auditing take on their highest values in the following period, leading to cheating and auditing activities in periods $4t+1, t = 1, ...$

Compare this to the case when $y = 0.4 > x_N$. Here, the amount of cheating is high enough to warrant auditing from the principal, who sets $\alpha = 1$ in periods $4t, t = 1, ...$ to coincide with those periods where 40% of the population is cheating. Thus in each cycle the principal audits all C players but also audits H players, for a cost of $0.4c_1 + 0.6c_2 = 2.2$. In the first scenario, with $y = 0.1$, the principal incurs a cost of $0.1c_3 + \delta c_1 = 0.1c_3 + 0.6c_1 = 1.6$.

Keeping the parameters of the problem fixed and just changing α_N to 0.15 as we did in the best response case, fictitious play still cycles using the threshold strategy, i.e., the behavior is not identical to the Best Response case. Moreover, cycling is the only behavior observed regardless of the discount factor, indicating that this may be analogous to the first simulation above where the threshold strategy was optimal regardless of discounting. Dropping α_N to just below 0.07 and δ to anything over 0.28, we get the over-audit strategy, as in Figure 4.5. This implies that the mechanism

by which fictitious play works might be similar to the best response case, with the difference being the expressions for the cut-off value for the different parameters.

Another way to obtain the over-audit strategy in fictitious play without changing α_N from the value above (0.15) is to change the values of c_1 and c_2 so that they are sufficiently close to each other. In the example above, keeping c_2 the same and increasing c_1 to be between 1.83 and less than 3, or decreasing c_2 to be higher than c_1 but less than 1.7, achieves the same result. This is not unexpected, since the threshold strategy alternates between periods of cost 0 and c_1 and the over-audit strategy has cost $c_2(\alpha_N + \epsilon)$ per period. As the Nash audit rate α_N and/or the cost c_2 associated with the over-audit strategy decreases and the discount factor is high enough, it becomes profitable to over-audit.

4.5 Comparison with Nash Equilibrium

Like the previous chapter, I point out some differences between the standard model of behavior and when agents learn according to the learning model described here.

Recall from Proposition 4.3.1 that the cost to the principal of playing the game in a fully rational setting is equal to $\frac{1}{1-\delta}\frac{c_2c_3}{c_3+c_2-c_1}$. Comparing this with the payoff from the threshold strategy in the best response case and assuming we start at the state $x = 1, \alpha = 1$, the cost is given by $c_1\frac{\delta}{1-\delta^2}$, which, when the constraint $0 < c_1 < c_2 < c_3$ is taken into account, is always less than $\frac{1}{1-\delta}\frac{c_2c_3}{c_3+c_2-c_1}$. However, when we compare the Nash payoff with the threshold strategy in the fictitious play case, the result of the comparison depends on the value of α_N. This is something that does not come into play in the Nash case since the payoff does not depend on α_N. Consequently, whether the principal is able to perform better in the fictitious play case depends on the value of α_N. If α_N is less than $\frac{1}{2}$ then the threshold strategy will perform better than the Nash equilibrium. This is because the strategy will do less auditing than in the best response case, which already has less cost than the Nash. When $\alpha_N > \frac{1}{2}$, then whether the threshold strategy performs better than the Nash or not depends on the parameters of the problem. For example, if α_N is close to 1, then the threshold strategy will perform better than the Nash if $c_2 \geq c_1^2 + \sqrt{c_1^4 - c_1^3}$.

4.6 Discussion

This chapter approached the problem presented in the last chapter from a different angle. The environment is assumed to be discrete and the agents, while still far from intelligent, have an added degree of sophistication — they try and think about how the principal is playing, even if they think of him as a stationary opponent. I examined the two extremes of strategies that depend on past play. The first is one when agents only react to what happened in the most recent period, and the second is when they take all past history into account.

I analytically showed that in the best response case, the principal always does better than the Nash solution and his preferred strategy is one of two: either to alternate periods of full auditing and periods of no auditing so that they induce and match cheating activity, or to constantly audit at a rate that is slightly higher than the Nash audit rate, thereby ensuring no cheating activity at the price of auditing a few honest agents. The latter strategy is selected by the principal when two things are in place: the Nash audit rate is quite low, leading to a not-so-significant portion of the population being bothered by the audit, and the discount factor is high enough, making the losses from auditing H players worth it in the long run.

While I do not derive the principal's optimal strategies for fictitious play, simulations suggest that the two strategies I examined in the best response case continue to be computationally optimal in a variety of settings under the fictitious play model. When $\alpha_N < \frac{1}{2}$, fictitious play does not have to do as much auditing as best response. This is not surprising as fictitious play seems to converge very quickly to α_N. It is also more sensitive to how the value of y is set, as can be seen in Figures 4.3 and 4.3. The same rationale for why the over-audit strategy is undominated seems to apply to fictitious play though, where again low levels of the Nash audit rate α_N and/or low values of c_2 suggest that it might be worth it to audit a few unfortunate H players in order to be able to acquire long-term gains.

Chapter 5

Conclusion

This thesis examined a variety of dynamic optimization problems in various settings and through the applications of different techniques. In some cases the dynamics of a problem can lead to difficulties in computing optimal policies and an overall loss of value for the principal, as was the case in the display advertising problem, where a one-shot, single-period version of the problem is much easier to solve than its multi-period counterpart. Conversely, dynamics allowed more interesting options for a principal when dealing with a learning population, where, depending on the learning mechanism used, the principal is able to extract higher payoffs in a game setting.

In the first part of the thesis, I showed how a publisher can deal with an uncertain supply situation in a display advertising problem in order to optimally fulfill a contract. The difficulties in this problem were mostly computational: the state space is simply too large to solve the problem. This is usually the biggest barrier in the face of optimal solution to dynamic problems. I showed that a special case of the problem admits a simple solution, and provided an approximation to the general case that allows the publisher to get around the computational difficulty.

The second and third parts of the thesis deal with a principal that wants to exploit learning in a population of agents to his advantage. The second chapter considers the case when agents learn according to replicator dynamics, and derives the principal's optimal policy under this model. The third chapter answers a similar question when agents play according to other learning models like best response and fictitious play.

There are several ways in which the work in this thesis can be extended. While the questions in

each of these chapters can be extended in their own right, it is most interesting to combine some of the ideas presented in this thesis. For example:

- The formulation of the ad delivery problem was very much influenced by problems of logistics and inventory management. There is not a lot of work that has been carried out at the marketing/operations interface (Karmarkar (1996)), but there are a lot of interesting questions. Most questions in inventory management problems revolve around fulfilling uncertain demand, but little or no work has been done to examine the effect of advertising on these policies. How can a firm manipulate social learning models through advertising in order to shape demand while also making capacity and pricing decisions? Can we characterize these joint inventory-pricing-advertising policies? How can one allocate budget optimally over the different dimensions? And how do the resulting policies change under various social learning models?

- The questions above consider a monopolist. What happens when firms compete *and* consumers learn? How do the policies change? And what do the equilibria of these games look like? What are the implications to standard industrial organization questions like entry barriers and price and product differentiation?

- In terms of computability, it seems reasonable to assume that complex integrated systems like the one in the second point above would not be amenable to the development of optimal *and* efficiently computable policies. For example, even when demand only depends on price, finding the optimal policies is computationally intractable.[1]. Are there policies that are both simple to implement and guarantee a solution that is close to optimal?

- How much does consumer rationality cost a firm? So far, Bayesian and behavioral learning agents have been considered in isolation. As I have shown in this thesis, it can benefit the principal when the population is behavioral, but how does this change when a fraction of the population is Bayesian? Can we quantify the losses to the firm —if any— as the degree of sophistication of the population increases? How does this affect the optimal policies for

[1] See *Coordinating inventory control and pricing strategies with random demand and fixed ordering cost* Chen and Simchi-Levi (2004).

problems where the population is either behavioral or rational? More interestingly, can we identify cases where the principal's payoffs depend on the fraction of the rational population in a non-monotonic fashion?

- The previous questions are all motivated by real-world business problems and concerns that a firm faces. One abstract question is understanding how advertising affects learning in a population and what learning models are more prone to be influenced by it. What are the best ways to implement public service campaigns that aim to increase social welfare through increasing awareness in a population?

Another related line of work is how to play a game against agents who are prone to making mistakes. In particular, can a principal exert some influence on the decisions made by a population through ways other than learning? One possible answer to this question mixes decision theory with the theory of computation. For example, can a firm combine economic models of thinking aversion[2] with computational complexity theory to provide a formal model of satisficing?[3] Recent experimental evidence in Caplin, Dean, and Martin (Forthcoming) suggests that subjects indeed satisfice when choosing from complex menus. How can such a model be used to design menus and choice sets that encourage certain biases that lead to maximizing a principal's objectives?

The two lines of inquiry in this thesis, dynamic problems and learning mechanisms, as well as areas in which both intersect, provide an exciting venue for research in the immediate future. With some of the classic work in economics already being ported into a dynamic format (e.g., Athey and Segal (2007) and Pavan, Segal, and Toikka (2009) for work on dynamic mechanism design), it is imperative that the learning aspects of these problems are infused into their dynamic analysis to provide an integrated framework that better describes questions in decision making and mechanism design, as well as questions in fields that rely on methods from these areas, like marketing and operations.

[2]For example, *The Price of Flexibility: Towards a Theory of Thinking Aversion.* Ortoleva (2008).

[3]See A Behavioral Model of Rational Choice. Simon (1955). Satisficing is when agents are unable to pick the optimal choice from a set and instead settle for something that is good enough.

Appendices

Appendix A

Proofs of Chapter 3

Proof of Proposition 3.2.2

Proof. Assumption 3.2.1 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. □

Proof of Theorem 3.3.2

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 3.1 and

are given by

$$\begin{aligned}
\dot{\alpha}(t) &= \alpha(t)\big[-(c_1x(t)+c_2(1-x(t)))-(c_3(1-\alpha(t))x(t)+c_1\alpha(t)x(t)+c_2\alpha(t)(1-x(t)))\big] \\
&= \alpha(t)(1-\alpha(t))[(c_3+c_2-c_1)x(t)-c_2) \qquad \text{(A.1)} \\
\dot{x}(t) &= x(t)\big[v_3(1-\alpha(t))-(v_1\alpha(t)(1-x(t))+v_2(1-\alpha(t))(1-x(t))+v_3(1-\alpha(t))x(t))\big] \\
&= x(t)(1-x(t))[(v_2-v_3-v_1)\alpha(t)+v_3-v_2]. \qquad \text{(A.2)}
\end{aligned}$$

Because of Assumptions 3.2.3 and 3.3.1, there are no equilibria on the boundary of the system described by (A.1) and (A.2). 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 (3.1) (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,

$$J=\begin{bmatrix} 0 & \frac{v1(v_3-v_2)}{(v_3+v_1-v_2)^2}(c_3+c_2-c_1) \\ \frac{c_2(c_3-c_1)}{(c_3+c_2-c_1)^2}(v_2-v_3-v_1) & 0 \end{bmatrix}$$

.

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 (A.1) and (A.2) (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 (A.1) and (A.2),

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

Rearranging terms and exponentiating, I end up with

$$B(x,\alpha) = \alpha^{v_3 - v_2}(1-\alpha)^{2(v_2 - v_3) - v_1}x^{-c_2}(1-x)^{c_1 - c_3}. \tag{A.3}$$

Now let (x, α) be a solution to the system (A.1) and (A.2). 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 A.0.1. *$\dot{B}(x, \alpha) = 0$ for any solution $(x(t), \alpha(t))$ to* (A.1) *and* (A.2)*.*

Proof. See below. □

Claim A.0.1 implies that the orbits described by (A.3) 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 (A.1) and (A.2). □

Proof of Claim A.0.1

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 provides

$$\frac{\partial}{\partial x}(\alpha^c(1-\alpha)^s x^a(1-x)^{-b}) = c(1-x)^{-b}x^{-a}(1-\alpha)^s\alpha^{c-1} - s(1-x)^{-b}x^{-a}(1-\alpha)^{s-1}\alpha^c. \quad \text{(A.4)}$$

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

$$\frac{\partial}{\partial \alpha}(\alpha^c(1-\alpha)^s x^a(1-x)^{-b}) = (1-x)^{-b-1}x^{-a-1}(a(-1+x)+bx)(1-\alpha)^s\alpha^c. \quad \text{(A.5)}$$

Using (A.1), (A.2), (A.4), and (A.5) 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 3.3.3

Proof. Consider a period of length T. We can rewrite (A.1) as

$$\begin{aligned}
\frac{\dot{\alpha}(t)}{\alpha(t)(1-\alpha(t))} &= ((c_1 - c_3 - c_2)x(t) + c_2) \\
\int_0^T \frac{\dot{\alpha}(t)}{\alpha(t)(1-\alpha(t))}dt &= \int_0^T ((c_1 - c_3 - c_2)x(t) + c_2)dt \\
\ln\frac{\alpha(t)}{1-\alpha(t)}\Big|_{t=0}^{t=T} &= c_2T - (c_1 - c_3 - c_2)\int_0^T x(t)dt.
\end{aligned}$$

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 (3.1). 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 (3.1). □

Proof of Lemma 3.4.1

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

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

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

$$\alpha^*(t) = \begin{cases} 0, & \lambda(t) < \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}; \\ 1, & \lambda(t) > \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}; \\ [0,1], & \lambda(t) = \frac{c_2 + (c_1 - c_2 - c_3)x(t)}{(p+q)x(t)(1-x(t))}. \end{cases} \tag{A.6}$$

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 3.4.2

Proof. Denoting the function inside the integral in (3.12) 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,

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

After some algebra and simplifying the above, I get

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

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

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

which is time-independent and a function of the parameters of the problem. □

Proof of Theorem 3.4.3

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 3.4.2, a necessary condition for the optimal path $x^*(t)$ to minimize (3.12) (and consequently (3.6)), 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 3.4.2. 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 (3.3), 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 3.2.3, $\bar{x}$ 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 (A.6), 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 3.4.1, 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 3.4.2, 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. This indicates 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 $x(t) < \bar{x}$. □

Proof of Theorem 3.5.1

Proof. Denote the audit rate in the behavioral setting by α_B. From Theorem 3.4.3, α_B is given by $\frac{q}{p+q}$. Replacing p and q by their values from Section 4.2.1, I find that the audit rate α_B is given by

$$\alpha_B = \frac{\frac{v_3 - v_1}{v_3}}{1 + \frac{v_3 - v_1}{v_3}} \tag{A.7}$$

which is always *strictly less* than the Nash audit rate in (3.1). Now consider the situation as $r \to 0$. In the fully rational setting this does not affect the outcome, and the principal and the population keep playing the strategies prescribed by (3.1) forever. Under the behavioral setting however, letting $r \to 0$ has a drastic effect on the cutoff value $\bar{x}$. From Equation (3.13) in Lemma 3.4.2,

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

Substituting for p and q, and taking the limit,

$$\lim_{r \to 0} \bar{x} = \frac{c_1 - c_2 + \frac{c_3(-v_2+v_3)}{v_3} - \sqrt{\left(c_1 - c_2 + \frac{c_3(-v_2+v_3)}{v_3}\right)^2}}{2\left(c_1 - c_2 + \frac{c_3(-v_2+v_3)}{v3}\right)} = 0.$$

which concludes the proof. □

Proof of Theorem 3.7.1

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:

$$\dot{x}(t) = x(t)(1-x(t))\big((1-\alpha(t))a_2 - \alpha(t)a_1\big) \tag{A.8}$$

$$\dot{\alpha}(t) = \alpha(t)(1-\alpha(t))\big((1-x(t))b_1 - x(t)b_2\big). \tag{A.9}$$

Let $x(0) = \frac{b_1+\epsilon}{b_1+b_2}$, for any $\epsilon \in (0, b_2)$, then from (A.9),

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

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 $\lim_{t\to\infty}(x(t), \alpha(t)) \to (1, 0)$. 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. □

Proof of Theorem 3.7.2

Proof. Using (A.9), 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

$$\min_{x(t)} \int_0^\infty e^{-rt}\left(\frac{a_2b_1(1-x(t)) + a_1b_2x(t) + \frac{(b_1-(b_1+b_2)x(t))\dot{x}(t)}{(1-x(t))x(t)}}{a_1+a_2}\right)dt \tag{A.10}$$

where $r > 0$ is a discount factor as before. Using the Euler-Lagrange equation to solve (A.10) leads to the following condition that should be satisfied by the optimal $x(t)$

$$\frac{e^{-rt}\left(-b_1 r + (a_1 b_2 e^{rt} - a_2 b_1)x(t)\right)}{(a_1 + a_2)x(t)} = 0. \tag{A.11}$$

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

$$x^*(t) = \frac{b_1 r}{a_1 b_2 e^{rt} - a_2 b_1}. \tag{A.12}$$

Then,

$$\lim_{r \to 0} x^*(t) = \lim_{r \to 0} \frac{b_1 r}{a_1 b_2 e^{rt} - a_2 b_1} = 0. \tag{A.13}$$

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

Proof of Proposition 3.8.1

Proof. The proof follows Section 3.4.2.1. 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 (3.7), 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

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

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

Proof of Proposition 3.8.2

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

$$\dot{x}(t) = x(t)(v_3(1 - \alpha(t)) - ((x(t) + \rho)v_3(1 - \alpha(t)) + (1 - x(t))v_1)). \tag{A.14}$$

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 (A.14), where

$$\bar{\alpha} = \frac{v_3 - v_1 + v_1(x_N - \rho) - v_3(x_N - \rho) - v_3\rho}{v_3(1 - x_N)}. \tag{A.15}$$

Obviously, $\bar{\alpha} < \alpha_N$ or else $p_c^* < 1$ (by Lemma 3.8.1), in which case the construction above fails. To verify that it is indeed the case, replace x_N by its value from (3.1) and subtract the Nash audit rate α_N from (A.15), leading to

$$\bar{\alpha} - \alpha_N = \frac{(c_3 + c_2 - c_1)v_1\rho}{(c_1 - c_3)v_3} \tag{A.16}$$

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. □

Proof of Proposition 3.8.3

Proof. This follows immediately from Lemma 3.8.1. 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 3.2.2. It follows that $\alpha^*(t) \geq \alpha_N$ and the rational population plays as in Lemma 3.8.1. $\square$

Appendix B

Code for Simulations

This appendix contains the MATLAB code necessary to run the simulations in Chapter 3 of this thesis. Working with a dynamic program, the state space has to be discretized. The first file, gridproj.m, is used throughout the code to project any values onto a discrete grid over which the state space is defined.

```
function [index]=gridproj(grid,point)
gridlength=length(grid);
gridl=1;
gridM=gridlength;
gridm=floor((gridl+gridM)/2);


if point>grid(gridlength-1)


    if 1-point<point-grid(gridlength-1)
        index=gridlength;
    else
        index=gridlength-1;
    end
elseif point<grid(2)
    if point<grid(2)-point
```

```
        index=1;
    else
        index=2;
    end
else
while not(grid(gridm)<=point && point < grid(gridm+1))
    if point<grid(gridm)
        gridM=gridm;
    else
        gridl=gridm;
    end
    gridm=floor((gridl+gridM)/2);
end
if point-grid(gridm)<grid(gridm+1)-point
    index=gridm;
else
    index=gridm+1;
end
end
```

The next function, Bellman2.m, solves the Bellman equation for a variety of learning mechanisms. Best response, fictitious play, truncated fictitious play, and weighted fictitious play are all included in the code.

```
function [Jvecout,polout]=Bellman2(grid,kappa,Jvec,C1,C2,C3,alpha_nash,y,t,T)


Jvecout=zeros(size(Jvec));
polout=zeros(size(Jvec));
threshold=alpha_nash;
```

```
k = 3;


for i=1:length(Jvec)
    h=grid(i);
    tempmin=10^10;
    tempalpha=-1;
    if h > threshold
        x=0;
    elseif h < threshold
        x=1;
    else
        x=y;
    end
    for alpha=grid
        nexth=(h*(T-t)+alpha)/(T-t+1);
%         nexth=alpha; %BR
%         nexth=(h*(k-1)+alpha)/k; % limited memory of k periods
%         nexth=(lambda * h*(T-t)+(1-lambda)*alpha)/(T-t+1);
        nextj=gridproj(grid,nexth);
        tempcost= x*C3+alpha*x*(C1-C2-C3)+alpha*C2+kappa*Jvec(nextj);
        if tempcost<tempmin
            tempmin=tempcost;
            tempalpha=alpha;
        end
    end


    Jvecout(i)=tempmin;
```

```
    polout(i)=tempalpha;
end
```

Next, script.m is the script file that contains the values for the parameters. It calls the main function, Bellman2.m and draws the graphs based on the outcome of the algorithm. The discount factor is called kappa, not delta, as in that part of the thesis.

```
steps=100;


grid=0:1/steps:1; % grid from 0 to 1 with distance 1/steps
gridl=length(grid);


C1=1;
C2=3;
C3=10;
alpha_nash=0.25;
y=0.1;
kappa=0.6;
lambda=0.5;
Jvec=zeros(gridl,1);
% Jvec=zeros(gridl,gridl);  %fictitious
T=30;
for t=1:T
    t
%     [Jvec,pol]=Bellman2(grid,kappa,Jvec,C1,C2,C3,y);    %BR
   [Jvec,pol]=Bellman2(grid,kappa,Jvec,C1,C2,C3,alpha_nash,y,t,T);  %fictitous
%      [Jvec,pol]=Bellman2(grid,kappa,lambda,Jvec,C1,C2,C3,alpha_nash,y,t,T);  %fictitous with ex
%     [Jvec,pol]=Bellman(grid,kappa,Jvec,C1,C2,C3,p,q);
%   [Jvec,pol]=Bellman(grid,kappa,Jvec,C1,C2,C3);
```

```
end


k=3;

 Htrajectories=zeros(gridl,T);

trajectories=zeros(gridl,T);

Xtrajectories=zeros(gridl,T);%

threshold = alpha_nash;

delta_threshold = alpha_nash*(C2+y*(C1-C2-C3))/(C1-alpha_nash*(C2+y*(C1-C2-C3)));

delta_th_2= alpha_nash * C2/(C1-alpha_nash*C2);

nash_cheating=C2/(C3+C2-C1);

for i=1:gridl

h=grid(i);

% x=grid(i);

trajectories(i,1)=pol(i);

Htrajectories(i,1)=h;

% Xtrajectories(i,1)=x;

alpha=pol(i);

for t=2:T

h=(h*(t-1)+alpha)/t;

% h=alpha; %BR

% h=(h*(k-1)+alpha)/k; %limited memory of k periods

% h=(lambda * h*(T-t)+(1-lambda)*alpha)/(T-t+1); %discounted past

if h > threshold

        x=0;

    elseif h < threshold

        x=1;

    else
```

```
        x=y;
    end
index=gridproj(grid,h);
alpha=pol(index);
trajectories(i,t)=alpha;
Htrajectories(i,t)=h;
Xtrajectories(i,t)=x;
end
end


plot(Htrajectories(1,:))
hold on;
plot(trajectories(1,:),'r')
hold on;
plot(Xtrajectories(1,:),'g')


%  string='';
%
% for k=1:20
%    data1=k;
%    data2=trajectories(56,k);
%    string=[string,'('];
%    string=[string,num2str(data1)];
%    string=[string,','];
%    string=[string,num2str(data2)];
%    string=[string,') --'];
% end
```

```
%
% string='';
%
% for k=1:20
%    data1=k;
%    data2=Xtrajectories(56,k);
%    string=[string,'('];
%    string=[string,num2str(data1)];
%    string=[string,','];
%    string=[string,num2str(data2)];
%    string=[string,') --'];
% end


% for k=1:100
%    data1=k;
%    data2=Jvec(k);
%    string=[string,'('];
%    string=[string,num2str(data1)];
%    string=[string,','];
%    string=[string,num2str(data2)];
%    string=[string,') --'];
% end


%  string
```

Finally, I include the code that simulates replicator dynamics, even though no simulation was included in this part of the thesis.

```
function [Jvecout,polout]=Bellman(grid,kappa,Jvec,C1,C2,C3,p,q,l,s)
```

```
Jvecout=zeros(size(Jvec));

polout=zeros(size(Jvec));

for i=1:length(Jvec)

    x=grid(i);

    tempmin=10^10;

    tempalpha=-1;

    for alpha=grid

%         nextx=x*(1+(1-x)*(1-2*alpha));

        nextx=x*(1+(1-x)*(p-alpha*(q+p)));

%         nextx=x*(1+(1-x)*(s*(1-alpha)^2-alpha*(p-q+alpha*(1-p+q))));

        nextj=gridproj(grid,nextx);

        tempcost= x*C3+alpha*x*(C1-C2-C3)+alpha*C2+kappa*Jvec(nextj);

        if tempcost<tempmin

            tempmin=tempcost;

            tempalpha=alpha;

        end

    end

    Jvecout(i)=tempmin;

    polout(i)=tempalpha;

end
```

Bibliography

Acemoglu, D., K. Bimpikis, and A. Ozdaglar (2009): "Communication Dynamics in Endogenous Social Networks," *Working Paper*.

Acemoglu, D., M. Dahleh, I. Lobel, and A. Ozdaglar (2008): "Bayesian learning in social networks," *National Bureau of Economic Research Working Paper*.

Athey, S., and I. Segal (2007): "An efficient dynamic mechanism," *Unpublished manuscript, Harvard University*.

Babaioff, M., J. Hartline, and R. Kleinberg (2008): "Selling banner ads: Online algorithms with buyback," in *Fourth Workshop on Ad Auctions*.

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.

Bergemann, D., and S. Morris (2005): "Robust mechanism design," *Econometrica*, pp. 1771–1813.

Bertsekas, D. (1995): "Dynamic programming and optimal control," *Athena Scientific Publishing*.

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.

Boutilier, C., D. Parkes, T. Sandholm, and W. Walsh (2008): "Expressive banner ad auctions and model-based online optimization for clearing," in *Proceedings of National Conference on Artificial Intelligence (AAAI)*.

Brown, G. (1951): "Iterative solution of games by fictitious play," *Activity Analysis of Production and Allocation*, 13(1), 374–376.

Butters, G. (1977): "Equilibrium distributions of sales and advertising prices," *The Review of Economic Studies*, 44(3), 465–491.

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*, Vol. 3, 587–704.

Caplin, A., M. Dean, and D. Martin (Forthcoming): "Search and Satisficing," *American Economic Review*.

Chen, X., and D. Simchi-Levi (2004): "Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The finite horizon case," *Operations Research*, pp. 887–896.

Cole, R., Y. Dodis, and T. Roughgarden (2006): "How much can taxes help selfish routing?," *Journal of Computer and System Sciences*, 72(3), 444–467.

Constantin, F., J. Feldman, S. Muthukrishnan, and M. Pál (2008): "Online ad slotting with cancellations," in *4th Workshop on Ad Auctions*.

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 crackdown on corruption in the city of Buenos Aires," *Journal of Law and Economics*, pp. 269–292.

Dorfman, R., and P. Steiner (1954): "Optimal advertising and optimal quality," *The American Economic Review*, 44(5), 826–836.

Eeckhout, J., N. Persico, and P. Todd (2010): "A Theory of Optimal Random Crackdowns," *American Economic Review*.

Feige, U., N. Immorlica, V. Mirrokni, and H. Nazerzadeh (2008): "A combinatorial allocation mechanism with penalties for banner advertising," in *Proceeding of the 17th international conference on World Wide Web*, pp. 169–178. ACM.

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. Association of Computing Machinery.

Fischer, S., and B. Vocking (2004): "On the evolution of selfish routing," *Algorithms–ESA 2004*, pp. 323–334.

Fudenberg, D., and D. Levine (1989): "Reputation and equilibrium selection in games with a patient player," *Econometrica: Journal of the Econometric Society*, pp. 759–778.

——— (1992): "Maintaining a reputation when strategies are imperfectly observed," *The Review of Economic Studies*, 59(3), 561–579.

——— (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.

Garey, M., and D. Johnson (1979): *Computers and intractability. A guide to the theory of NP-completeness. A Series of Books in the Mathematical Sciences*. WH Freeman and Company, San Francisco, CA.

Ghosh, A., P. McAfee, K. Papineni, and S. Vassilvitskii (2009): "Bidding for representative allocations for display advertising," *Internet and Network Economics*, pp. 208–219.

Golub, B., and M. Jackson (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2(1), 112–149.

Grether, D. (1992): "Testing Bayes rule and the representativeness heuristic: Some experimental evidence," *Journal of Economic Behavior & Organization*, 17(1), 31–57.

Griffin, D., and A. Tversky (1992): "The weighing of evidence and the determinants of confidence* 1," *Cognitive Psychology*, 24(3), 411–435.

Halman, N., D. Klabjan, M. Mostagir, J. Orlin, and D. Simchi-Levi (2009): "A Fully Polynomial-Time Approximation Scheme for Single-Item Stochastic Inventory Control with Discrete Demand," *Mathematics of Operations Research*, 34(3), 674–685.

Holt, C., and A. Smith (2009): "An update on Bayesian updating," *Journal of Economic Behavior & Organization*, 69(2), 125–134.

Jackson, M. (2008): *Social and economic networks*. Princeton University Press.

Kahneman, D., A. Tversky, Decisions, and Designs (1977): *Intuitive prediction: Biases and corrective procedures*. Defense Technical Information Center.

Kamenica, E., and M. Gentzkow (2009): "Bayesian persuasion," *National Bureau of Economic Research Working Paper*.

Kamien, M., and N. Schwartz (1980): *Dynamic optimization: the calculus of variations and optimal control in economics and management*. North Holland Amsterdam.

Karmarkar, U. (1996): "Integrative research in marketing and operations management," *Journal of Marketing Research*, 33(2), 125–133.

Krajbich, I., C. Camerer, J. Ledyard, and A. Rangel (2009): "Using Neural Measures of Economic Value to Solve the Public Goods Free-Rider Problem," *Science*, 326(5952), 596.

Levi, R., M. Pal, R. Roundy, and D. Shmoys (2007): "Approximation algorithms for stochastic inventory control models," *Mathematics of Operations Research*, 32(2), 284.

Lim, M., R. Metzler, and Y. Bar-Yam (2007): "Global pattern formation and ethnic/cultural violence," *Science*, 317(5844), 1540.

Lui, F. (1986): "A dynamic model of corruption deterrence," *Journal of Public Economics*, 31(2), 215–236.

Mostagir, M. (2005): "Fully polynomial time approximation schemes for sequential decision problems," Master's thesis, Massachusetts Institute of Technology.

——— (2010a): "Exploiting Myopic Learning," Working paper.

——— (2010b): "Optimal Delivery in Display Advertising," Working paper.

Myerson, R. (1981): "Optimal auction design," *Mathematics of operations research*, 6(1), 58.

Myerson, R. B. (1988): "Mechanism Design," Discussion Papers 796, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

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," *Working Paper*.

Pavan, A., I. Segal, and J. Toikka (2008): "Dynamic mechanism design: Revenue equivalence, profit maximization and information disclosure," *Northwestern University, unpublished working paper*.

——— (2009): "Dynamic mechanism design: Incentive compatibility, profit maximization and information disclosure," .

Perko, L. (2001): *Differential equations and dynamical systems*. Springer Verlag.

Rabin, M. (1998): "Psychology and economics," *Journal of Economic Literature*, 36(1), 11–46.

Roughgarden, T. (2006): "On the severity of Braess's paradox: designing networks for selfish users is hard," *Journal of Computer and System Sciences*, 72(5), 922–953.

——— (2007): "Routing games," *Algorithmic Game Theory*, p. 461.

Roughgarden, T., and É. Tardos (2002): "How bad is selfish routing?," *Journal of the Association of Computing Machinery (JACM)*, 49(2), 259.

Simon, H. (1955): "A behavioral model of rational choice," *The Quarterly Journal of Economics*, 69(1), 99.

Smith, M. (1979): "The existence, uniqueness and stability of traffic equilibria," *Transportation Research Part B: Methodological*, 13(4), 295–304.

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.

Woeginger, G. (1999): "When does a dynamic programming formulation guarantee the existence of an FPTAS?," in *Proceedings of the Tenth Annual Symposium on Discrete Algorithms*, pp. 820–829. Society for Industrial and Applied Mathematics Philadelphia, PA, USA.

Yano, C., and H. Lee (1995): "Lot sizing with random yields: a review," *Operations Research*, pp. 311–334.

Young, H. (1993): "The evolution of conventions," *Econometrica: Journal of the Econometric Society*, 61(1), 57–84.

Zipkin, P. (2000): "Foundations of inventory management," *McGraw-Hill Boston*.