MACHINE LEARNING FOR STRATEGIC INFERENCE

IN-KOO CHO AND JONATHAN LIBGOBER

Abstract. We introduce a framework to study the interaction between a strategic players and an algorithm designer which guides the behavior of other parties. The algorithm is limited in the set of decision rules that can be used to prescribe actions. Nevertheless, appealing to the endogeneity of the data generating process, we show it is possible to guide behavior across a rich set of possible environments using limited details. More precisely, we show how *Adaptive Boosting* algorithms can be specified to induce behavior that is (approximately) as-if rational using only observed data, and in particular without needing to infer the optimal response from Bayes rule. The key property ensuring our algorithms can succeed is *weak learnability*, which we show holds as long as it is possible to find a best-fitting single-threshold classifier. The as-if rational benchmark is the optimal target in the absence of domain specific-knowledge, but even this may be unachievable without considering the algorithm design problem. We describe how our analysis provides a statistical perspective to the study of endogenous model misspecification.

1. Introduction

Algorithms play a significantly larger role in guiding economic behavior today than even a decade ago, and yet their importance is likely only going to grow further in the years to come. On the other hand, a number of economic phenomena appear to rely crucially on the presence of rational individuals on both sides of the relevant interaction. Of course, a number of papers have studied the implications of departures from the rational benchmark, and economists largely recognize that these assumptions of rationality have empirical implications. What is left open is whether algorithms are susceptible to the same kinds of biases, and if so, when.

This paper introduces a framework to study this question. In our model, a rational, strategic player (who we refer to as a *principal*) chooses a strategy when interacting with an algorithm that will prescribe actions to a stream of short lived actors (who we refer to as *agents*). The particular interaction we are interested in is one of *strategic inference*: We assume that the principal commits to a strategy which maps states into actions, and a rational agent would therefore need to update beliefs according to Bayes rule in response to the observation from the principal in order to act optimally. In contrast, the algorithm has access to observations on what happened in previous interactions. Our question is how

Date: September 22, 2020.

We thank Juan Carrillo, Grigory Franguridi, Roger Moon, Xiaosheng Mu, Guofu Tan, Joel Sobel and Erik Strand for helpful conversations and comments, and seminar audiences at AMETS, the NSF/NBER/CEME Conference on Mathematical Economics, Rochester, UC-San Diego, and USC. This project started when the first author was visiting the University of Southern California. We are grateful for hopitality and support from USC. Financial support from the National Science Foundation is gratefully acknowledged.

the outcome compares to the case where the agent is indeed rational. Notice that if the agent's behavior departs from rationality, it may be both that their actions are suboptimal, as well as that the principal seeks to exploit this when choosing their strategy.

In our model, a significant limitation of the algorithm is that the set of decision rules that can be used is severely limited. Though perhaps unusual in the economic theory literature, we believe this aspect is well motivated from practical considerations. In typical machine learning problem, a simple prediction (for instance, a "yes-no" recommendation) is sought for an observation among a very large set of possibilities. Seeking to find the correct recommendation for each one may be intractable (or even undesirable for overfitting reasons), and so a simpler set may be used as a baseline. On the other hand, it may still be possible to construct a new decision rule if the algorithm specifies how this should be done in advance. In our model, this takes the form of assuming the algorithm is limited in what can be fit to the data, but is otherwise flexible, in a way we will make precise below.

While this explicit assumption is perhaps unusual in economic theory, it is not too different from the assumption that a decisionmaker may use a misspecified model when learning about their environment. The difference is that the algorithm in our framework is concerned explicitly with *prediction*. In the (currently very active) literature on model misspecification (see, for instance, Esponda and Pouzo (2014)), a decisionmaker is assumed to be potentially incorrect regarding the set of possible parameters, but otherwise uses an optimally chosen decision rule. In contrast, in our framework, this decision rule is restricted. In other words, we are not (directly) interested in learning how to make an optimal *inference*, but rather an optimal *prediction*. That said, perhaps more interestingly, the fact that our algorithm can construct new decision rules from old ones in our model means that the degree of misspecification in the decision rules used is *endogenously chosen*.

With these observations aside, the natural question is what one should expect to happen given limitations in what kinds of reactions from the agent can be specified by the algorithm. There are two competing forces at play here. Perhaps more natural is that an algorithm designer may allow them to have greater commitment than a rational agent would have by limiting the set of decision rules the agent can use. This would suggest that the limitation in the set of decision rules could be helpful for the agent. More subtle is the competing force which would lead the seller to exploit the algorithm. For our purposes, the key force at play was identified by Rubinstein (1993), which illustrated how limiting the set of decision rules that can be utilized provides scope for exploitation. This paper showed that if a *rational* decisionmaker is *restricted* to use a binary threshold classifier—i.e., one that makes the same decision on a given side of a fixed threshold—then the seller can price discriminate by utilizing a particular form of randomization which "fools" these buyers into making a decision which is suboptimal, given the realized price.[1] Our framework nests Rubinstein (1993) as a special case, but considers more general environments

[1]The reasoning behind this result is as follows. First, the optimally chosen classifier chosen can do strictly better than simply randomizing the guess, implying that the seller can exploit the incentives of the buyer in order to manipulate the decision rule. On the other hand, it is impossible for threshold rules to implement the optimal decision with probability 1 when this rational rule is non-monotone in the price. The first point implies the buyer trades off against errors, and the second point implies that the tradeoff falls short of the fully rational response. As a result, the seller can force a different decision than would be rationally optimal for these buyers (with arbitrarily high probability).

as well. Indeed, it is straightforward to verify that the same strategies he uses can exploit a buyer limited in the models that can prescribe behavior.

The discussion suggests that in some environments, it may be that using the restricted rules may be helpful for the algorithm, but that in other environments, it may be that the restriction can be exploited. Our motivation is to have the algorithm recommendations be as data-driven as possible, and therefore not seek to utilize some added understanding of the environment to figure out which case is which. We therefore try to find an algorithm that does well uniformly across all possible environments, in a detail-free way. Of course, we can then think about the outcomes that would emerge in any particular environment. In terms of the objective of the algorithm designer, however, the goal is to do as well as possible in all possible environments—in the case of a seller on Amazon or Etsy, for instance, to ensure that there is no particular product being sold where the algorithm yields bad recommendations.

Our analysis elucidates a tension between the ability to fit *rich* and *coarse* sets of models. A consideration of Rubinstein (1993) is instructive to appreciate this point. His work shows that, if a decisionmaker is limited in the decision rules that can be utilized, then there is a potential for exploitation. In order to combat this temptation, one may seek to add more models to be fit to the data; in other words, to make the set of models richer. Indeed, a decisionmaker could prevent the particular instance of exploitation he highlights by adding additional models to the data. However, adding richer models to the data may have other undesirable consequences. This point is made clear by an appeal to the machine learning literature; finding the best fitting model within a set of models may be computationally demanding if this set is very large.

Our proposed solution is to use the Adaptive Boosting algorithm (Schapire and Freund (2012)), which specifies exactly how to construct a decision rule as a weighted combination of classifiers, with the weights specified by the algorithm. The algorithm requires us to be able to (repeatedly) fit a classifier to some distribution over prices and outcomes, from some set of baseline classifiers.

Returning to the particular setting at hand, to ensure that this algorithm yields rational responses, the requirement on the set of model (called *weak learnability*) that can be fit to data is significantly less demanding. We provide results which show how to check it in several straightforward applications, sometimes using a dramatically smaller set of classifiers than one might need otherwise. Returning to the discussion of Rubinstein (1993), we see that the issue with single threshold classifiers is that they are not *strong learners* (i.e., they cannot ensure the optimal decision is taken with probability 1 following any price), even though they are *weak learners* (i.e., they can outperform random guesses when chosen optimally). The remarkable property of the Adaptive Boosting algorithm is that it shows weak learnability is sufficient to construct a classifier that yields a similar guarantee as under a model class satisfying strong learnability. It is interesting that part of the intuition for the main result in Rubinstein (1993)—which relies upon the buyer being able to strictly improve payoffs beyond a trivial default—exactly tells us how to overcome the main conclusion, once we have the algorithm in hand.

At first glance, it appears that there is a significant gap between the weak learnability requirement and rationality. Rationality requires, in principle, very rich decision rules to be used, and for the performance of them to leave very little room for error. Weak

learnability does not, and only requires a uniform improvement over a random guess. It is therefore perhaps surprising that in our exercise, the turns out to be no gap at all. Due to weak learnability, the apparent gap in rationality caused by the limitation in the decision rules that can be fit to data can be overcome by a clever choice of algorithm. The result is that the algorithm can induce rational behavior without knowing anything beyond the observed data from past interactions. In contrast, *strong* learnability (i.e., prescribing the optimal action with high probability) may very well require precise knowledge of the principal's strategy.

We briefly mention that the algorithm design we study accommodates a rich possible action space, even with the *same* restrictions in what can be fit to data. This is in sharp contrast to other papers in the large literature on "decisionmakers as statisticians" (reviewed below), which use similar motivation to study departures from rationality. As discussed below, these papers have typically focused on the binary action case. This limitation is very natural—many of the key results from machine learning which arise when there are two possible predictions do not extend easily (or even at all) to the case of multiple actions. However, we can handle this in our problem, suggesting our algorithm is of broader interest.

Our hopeful contribution is in formalizing a way of applying machine learning methods to answer new questions relevant to microeconomic theorists, and visa versa. While strategic inference settings have been widely studied in economics since Akerlof, we are not aware of these applications having been addressed by the statistics literature. Our model is deliberately abstract, in order to provide general principles on when the problem of model misspecification can be overcome. Our message is that while it is not possible to guarantee that rationality emerges for arbitrarily data generating process, it *is* possible if the data generating process is endogenous to the statistical algorithm. This argument requires some additional steps using incentives of the actors to demonstrates that the resulting output does in fact correspond to what is traditionally thought of as subgame perfection. The endogeneity issue makes the problem no longer a pure statistical exercise. The modifications our analysis requires extend beyond the initial need to show that it is possible to do better than random guessing in this environment. As our analysis elucidates, AdaBoost is capable of handling a *particular* kind of unboundedness in the cardinality of the action space. It is thus necessary to discipline the environment further in order to achieve our results.

2. Literature

This paper is most closely related to the literature on learning in games when players' behavior depends on a statistical method. The single-agent problem is a particular special case. Single agent versions of this problem are the focus of Al-Najjar (2009) and Al-Najjar and Pai (2014). However, it is worth emphasizing that the data buyers receive is *endogenous* in our setting because of the strategic interactions. In contrast, their benchmarks correspond to the case of exogenous data. This problem is also studied in Spiegler (2016), who focuses on causality and defines a solution concept for behavior that arises from individuals fitting a directed acyclic graph to past observations. More recently, Zhao, Ke, Wang, and Hsieh (2020) take a decision-theoretic approach in a single-agent setting with

lotteries, showing how a relaxation of the independence axiom leads to a neural-network representation of preferences.

Taking these approaches to games, the literature has still for the most part focused on settings where the interactions between players is *static*, typically imposing finiteness to a degree that rules out the game of Rubinstein (1993). In contrast, our setting is a simple, two-player (and two-move) sequential game. Cherry and Salant (2019) discuss a procedure whereby players' behavior arises from a statistical rule estimated by sampling past actions. This leads to an endogeneity issue similar to the one present in our environment, i.e., an interaction between the data generating process and the statistical method used to evaluate it. Eliaz and Spiegler (2018) study the problem of a statistician estimating a model in order to help an agent take an action, motivated (like us) by issues involved with the interaction between rational plays and statistical algorithms. Liang (2018), like us, is focused on games of incomplete information, asking when a class of learning rules leads to rationalizable behavior. Focusing on the application of model selection in econometrics, Olea, Ortoleva, Pai, and Prat (2019) study an auction model and ask which statistical models achieve the highest confidence in results as a function of a particular dataset.

On the other hand, the literature on learning in extensive form games has typically assumed that agents experiment optimally, and hence embeds notion of rationality on the part of agents which we dispense with in this paper. Classic contributions include Fudenberg and Kreps (1995), Fudenberg and Levine (1993) and Fudenberg and Levine (2006). Most of this literature has focused on cases where there is no exogenous uncertainty regarding a player's type, and asking whether self-confirming behavior emerges as the outcome. An important exception is Fudenberg and He (2018), who study the steady-state outcomes from experimentation in a signalling game. While a rational agent in our game would need to form an expectation over an exogenous random variable, signalling issues do not arise because our seller has commitment.

Perhaps closest in motivation is the computer science literature studying how well algorithms perform in strategic situations, as well as how rational actors (particularly sellers) may respond when facing them. Braverman, Mao, Schneider, and Weinberg (2018) consider optimal pricing of a seller repeatedly selling to a single buyer who repeatedly uses a no-regret learning algorithm. They show that, on the one hand, while a particular class of learning algorithms (i.e., those that are *mean-based*) are susceptible to exploitation, others would lead to the seller's optimal strategy simply being to use the Myersonian optimum. Deng, Schneider, and Sivan (2019) also study strategies against no-regret learners in a broad class of games without uncertainty, and consider whether a strategic player can guarantee a higher payoff than what would be implied by first-mover advantage. Blum, Hajiaghayi, Ligett, and Roth (2008) consider the Price of Anarchy (i.e., the ratio between first-best welfare and worst-case equilibrium welfare), and show in a broad class of games that this quantity is the same whether players use Nash strategies or regret-minimizing ones. Nekipelov, Syrgkanis, and Tardos (2015) assume players in a repeated auction use a no-regret learning algorithm, making similar behavioral assumptions as this paper. Their interest is in inferring the set of rationalizable actions from data.

While our motivation is very similar—and indeed, we seek to incorporate several aspects of this literature's conceptual framework—there are several notable differences. First, this literature typically assumes particular algorithms or objectives (e.g., no regret learning)

which differ from standard Bayesian rationality. In contrast, our results illustrate how different constraints on algorithms (namely, whether or not combining classifiers is feasible) may lead to different outcomes emerging as possible. Second, our general framework incorporates a different set of single-agent applications which extend beyond particular pricing settings, where most (though admittedly not all) of this literature has focused. In particular, we are not aware of Lemons markets settings (such as Rubinstein (1993), for example) having been studied in this literature, which form our primary starting point. As a result, new technical issues (e.g., dealing with residual uncertainty in the correct actions) is not addressed in these papers to our knowledge. Third, we focus on jointly relating the incentives of the rational player and the algorithm's ability to approximate rationality. In contrast, the papers we are aware of typically study each separately, either asking how a rational player would maximize payoff against an algorithm, or comparing the performance of no-regret algorithms to fully optimal or Bayesian ones. Despite these differences, it appears that these results have not been fully appreciated in economics, and thus our hope is that this paper inspires further connection between the economics literature on decisionmakers as statisticians and the computer science literature on strategic choices against classes of algorithms.

3. Stage Games

We first describe the stage game interaction that the algorithm designer seeks to prescribe actions for on behalf of myopic actors (who may be, for instance, receivers, buyers, or agents, depending on the particular setting of interest). The stage game features a strategic actor as well. That said, our exposition in this section does not address how this actor's strategy and the algorithm are determined. This is done in Section 4, which describes the interaction which yields the algorithm and the strategic player's strategy and the relevant objective for each.

3.1. **Actions and Parameters.**

The timing of a stage game is similar to a principal agent model with principal private information (as in Maskin and Tirole (1992), for instance). Let Θ be the set of types endowed with a prior distribution π which is a common knowledge among players. This type is payoff relevant to both the principal and the agent. Define $\pi(\theta)$ as the probability that type $\theta \in \Theta$ is realized. Conditioned on the realized value of $\theta \in \Theta$, the principal takes an action $p \in \mathcal{P} \subset \mathbb{R}^n$ where $\mathcal{P}$ is a compact subset of $\mathbb{R}^n$. We emphasize that the distribution over principal's actions is determined in the algorithm game described in Section 4. Conditioned on p (but not θ), the agent chooses $a \in A$.

A strategy of the principal is

$$\sigma : \Theta \to \Delta(\mathcal{P})$$

where $\Delta(\cdot)$ is the probability distribution over the set. This strategy is determined in the Algorithm Game described in Section 4. Given this strategy, the stage game proceeds as follows:

S_1. An exogenous state $\theta \in \Theta$ is realized according to π, and
S_2. The principal's action $p \in \mathcal{P}$ is realized according to $\sigma(\cdot \mid \theta)$.
S_3. The agent takes action $a \in A$.

The stage game is a sender-receiver game in which an informed sender makes the first move. If we interpret $p = (p_1, \ldots, p_n)$ as a contract, and $a \in A = \{-1, 1\}$ as "reject" (-1) or "accept" (1), the stage game is a model of the informed principal (Maskin and Tirole (1992)). If p is interpreted as a message sent by the principal, and $a \in A$ as the price the agent pays, then the stage game becomes a signaling game (Spence (1973)).

3.2. **Payoffs and the Rational Benchmark.** We allow the stage game interaction to feature some additional payoff uncertainty observed by the strategic player, denoting this by $\kappa \in \mathcal{K}$. We denote the strategic player's payoff by $u(\theta, p, a, \kappa)$, and the agent's payoff by $v(\theta, p, a, \kappa)$, which we take to be (weakly) decreasing coordinatewise in p. Note that if the agent is (sequentially) rational, then his optimization problem is

$$\max_{a \in A} v(\theta, p, a, \kappa)\pi(\theta|p)$$

where $\pi(\theta|p)$ is the posterior probability assigned to θ conditioned on p by Bayes rule:

$$\pi(\theta|p) = \frac{\sigma(p|\theta)\pi(\theta)}{\sum_{\theta'} \sigma(p|\theta')\pi(\theta')}.$$

We define the *rational label*:

$$y^R : \Sigma \times \mathcal{P} \to A$$

to be a solution to the following optimization problem:

$$v(\theta, p, y^R(\sigma, p), \kappa)\pi(\theta|p) \geq v(\theta, p, a, \kappa)\pi(\theta|p) \qquad \forall a \in A$$

where $\pi(\theta|p)$ is computed via Bayes rule whenever $\sum_\theta \sigma(p|\theta)\pi(\theta) > 0$.[2]

Define σ^R as a best response of the principal against a Bayesian rational agent with perfect foresight:

$$\sum_{\theta,p,a} u(\theta, p, a, \kappa)\sigma^R(p|\theta)y^R(\sigma^R, p)\pi(\theta) \geq \sum_{\theta,p,a} u(\theta, p, a, \kappa)\sigma(p|\theta)y^R(\sigma^R, p)\pi(\theta) \qquad \forall \sigma.$$

By the construction, (σ^R, y^R) constitutes a perfect Bayesian equilibrium.[3]

3.2.1. *Rational Fringe.* For some applications, one could imagine the principal interacts with some agents who are algorithmic, and others who are rational. Indeed, the setting of Rubinstein (1993) features such a dichotomy, as we discuss.

We accommodate this as follows: Specifically, let r denote the probability that the agent's action is informed by the choices during the algorithm game. With probability $1 - r$, the buyer chooses action $y^R(\sigma, p)$. Let $i = 1$ denote the event that the buyer is *algorithmic* (that is, uses a strategy informed by the algorithm), and $i = 2$ denote the event that the buyer is *rational*. In this case, we allow the seller's profit to potentially depend on this type of the buyer, and in this case we abuse notation and add a fifth argument to the principal's utility, $u(\theta, p, a, \kappa, i)$.

Modelling the rational fringe allows us to borrow the existing intuition to provide one reason such incentives to exploit algorithms might emerge. When the seller's payoff depends on i, it may be more transparent to see why the conditions for exploitation we

[2]For a fixed σ, $y^R(\sigma, \cdot) : \mathcal{P} \to A$ is a strategy of the agent, satisfying sequential rationality.

[3]It is important to note that the principal optimize against the equilibrium strategy $y^R(\sigma^R, \cdot)$ of the agent, rather than $y^R(\sigma, \cdot)$.

present below may hold. See Footnote 4 for one explanation for how such a dependence might emerge, in large part motivated by Rubinstein (1993). From a theoretical perspective, however, we see no issue with simply "integrating out" the rational fringe, though admittedly it may be more work to write the principal's payoff as a function of primitives if doing so.

3.3. **Examples of Stage Games.** Before proceeding to the description of the algorithm game, we describe a few of the stage game interactions that are of primary interest. We will return to these later in order to illustrate the incentives for each party to potentially exploit the other. We note that the dependence on κ will be suppressed for these examples; its role is clarified in the description of the Algorithm Game.

3.3.1. *Rubinstein [1993].* While Rubinstein (1993) differs expositionally, we review the key ideas and describe how it falls under our framework. Suppose $\theta \in \Theta = \{L, H\}$. v_θ is the marginal utility of the good where $v_H > v_L > 0$. The prior probability distribution is $\pi(H) = \pi(L) = \frac{1}{2}$. The seller choose a price $p \in [v_L, v_H] \subset \mathcal{P} \subset \mathbb{R}$, conditioned on $\theta \in \Theta$. The action of a buyer is $a \in A = \{-1, 1\}$. A buyer responds to $p \in \mathcal{P}$ by purchasing $(a = 1)$ or not purchasing $(a = -1)$ the good at p.

The seller is facing a unit mass of infinitesimal buyers, who can be either type 1 or type 2. The proportion of type 1 buyer is $r \in (0, 1)$. The buyers differ in terms of the cost of sales. If $\theta = L$, the product costs c_L for the seller regardless of the types of the buyer. If $\theta = H$, the product costs c_i to serve type i buyer $(i \in \{1, 2\})$. We assume

$$c_1 > v_H > c_2 > v_L > c_L \tag{3.1}$$

$$rc_1 + (1 - r)c_2 > v_H \tag{3.2}$$

so that the agent is exposed to the lemon's problem. A buyer generates utility only if he purchases the good, whose payoff function is

$$v(\theta, p, a) = \begin{cases} 0 & \text{if } \ a = -1 \\ v_\theta - p & \text{if } \ a = 1. \end{cases}$$

The payoff of the seller[4] is

$$u(\theta, p, a, i) = \begin{cases} 0 & \text{if } \ a = -1 \\ p - c_L & \text{if } \ a = 1, \ \theta = L \\ p - c_i & \text{if } \ a = 1, \ \theta = H. \end{cases}$$

The unique Nash equilibrium strategy of the seller is

$$\sigma^R(\theta) = \begin{cases} v_L & \text{if } \ \theta = L \\ v_H & \text{if } \ \theta = H. \end{cases}$$

[4] Importantly, this payoff function features dependence of the principal's utility on whether the buyer is algorithmic or rational. For instance, suppose a seller seeks to sell a product both online, as well as in person, using the same terms in order to deter third parties from seeking to resell their product it online. This may require the retailer themselves offering the product, in order to prevent resalers from showing up first. It still may be the case, however, that the seller would lose money were the consumer to purchase.

The buyer's equilibrium strategy is

$$y^R(p) = \begin{cases} 1 & \text{if } p \leq v_L \\ -1 & \text{if } p > v_L. \end{cases}$$

The trading occurs only if $\theta = L$, and therefore, the equilibrium is inefficient. Note that the construction of y^R requires a precise information about v_L.

3.3.2. *Insurance.* The following is borrowed from Maskin and Tirole (1992). Suppose that the principal is a shipping company seeking to purchase insurance from an insurance company, an agent that is seeking to delegate the decision of whether to offer the terms put forth by the shipping company. The principal seeks insurance every period, but faces risk (e.g., due to the location of shipping demand is) that is idiosyncratic every period.

In this case, we imagine the principal choose terms within some bounded compact set $\mathcal{P} \subset \mathbb{R}^2$, where $p = (x, q)$ denotes a policy which provides a payment x in the event of a loss, and costs an amount q. If θ denotes the probability of a loss, then the principal's utility is:

$$u(\theta, p, a) = \begin{cases} (1-\theta)f(I-q) + \theta f(I-q-L+x) & a = 1 \\ (1-\theta)f(I) + \theta f(I-L) & a = -1 \end{cases},$$

for some concave f. The agent's utility is:

$$v(\theta, p, a) = \begin{cases} q - \theta x & a = 1 \\ 0 & a = -1 \end{cases}$$

It is natural to consider $\mathcal{P}$ whereby, against a rational buyer, the principal would seeks a high level of insurance when risk is high (i.e., $\theta = H$), and avoid insurance when risk is low (i.e., $\theta = L$). In contrast, the agent's payoff may be decreasing in the quantity of insurance when $\theta = H$, while increasing in the quantity of insurance when $\theta = L$.

3.3.3. *Spence [1973].* Our framework is general and can be expanded to cover other settings as well. Let us consider a labor market signaling model. Here, the "agent" takes the role of the firm and the "principal" takes the role of the worker from the Spence signalling model (as in, for instance, Maskin and Tirole (1992)). The true state is the productivity of the worker $\theta \in \Theta = \{H, L\}$ where $\pi(H) = \pi(L) = \frac{1}{2}$: $H > L$. Conditioned on θ, a worker chooses p which we interpret as education level. His strategy is

$$\sigma : \Theta \to \mathcal{P} \subset \mathbb{R}_+.$$

The payoff function of the principal is

$$u(\theta, p, a) = a - \frac{p}{\theta + 1}$$

We abstract away the competition among multiple firms in the labor market. Conditioned on p, the labor market wage is determined according to the expected productivity $\mathbf{E}(\theta : p)$ conditioned on p. The agent has to pay the principal the equal amount of the expected productivity because of (un-modeled) competition among firms. The agent's goal is to

make an accurate forecast about the expected productivity of the worker. The payoff of the agent is

$$v(\theta, p, a) = -(\theta - a)^2$$

If the support of $\sigma(p : H)$ is disjoint from the support of $\sigma(p : L)$, σ is a separating strategy. If a separating strategy is an equilibrium strategy, then the equilibrium is called a separating equilibrium. We often focus on the Riley outcome, which maximizes the ex ante expected payoff of the principal among all separating equilibria.

3.4. **Introducing Time.** Our question of interest is whether the agent can learn rational label y^R, if the stage game is repeated over time. As an intermediate step toward defining algorithm games, we describe our approach and assumptions involved with this step. In the next section, we discuss the algorithm choice that occurs on top of this.

By *expanded stage game*, we refer to a repetition of the stage game interaction, played over discrete time $t = 1, 2, \ldots$, where the stage game interaction occurs at every $t \geq 1$. Our substantive assumption is that (θ, p) is drawn IID across periods (whereas κ is fixed throughout all time). In this case, given (σ, σ_b), the expected payoff of the principal is[5]

$$\lim_{T\to\infty} \mathbf{E}\frac{1}{T}\sum_{t=1}^{T} u(\theta, p, a)\pi(p)\sigma(p|\theta)\sigma_{b,t}(a|p) \tag{3.3}$$

and the expected payoff of the agent is

$$\lim_{T\to\infty} \mathbf{E}\frac{1}{T}\sum_{t=1}^{T} v(\theta, p, a)\pi(p)\sigma(p|\theta)\sigma_{b,t}(a|p). \tag{3.4}$$

Note that a rational player chooses $\sigma_{b,t}(a \mid p) = y^R(\sigma, p)$ (for all t, in which case the limits drop out).

4. Algorithm game

Having outlined the basic timing of moves that occur and determine payoffs, we now describe the supergame which determines the agent's strategy. We refer to this as an *algorithm* game.

4.1. **Choices of Algorithms.** We will refer to the strategy an agent uses—which is output by the algorithm at every time—as a *classifier*, in line with the machine learning and computer science literature:

Definition 4.1. *A classifier is a function*

$$\gamma : P \to A.$$

This may additionally be referred to as either a strategy or a forecasting rule.

[5]Prior versions of this paper considered the case where $\delta < 1$; the main lessons remain valid for δ sufficiently large, although there are some added technical difficulties in the analysis of Section 7.4 this introduces.

In order to construct the classifier, the algorithm faces some computational constraints. More precisely, we assume that there is a fixed set of classifiers $\mathcal{H}$ (referred to as the *hypothesis class*) for which the algorithm can solve the following problem:

$$\min_{h\in\mathcal{H}} \sum_{p} \mathbf{1}[h(p) = y(p)]L(p), \tag{4.5}$$

for an arbitrary function L and function $y : \mathcal{P} \to A$. We refer to this step as finding the *best fitting hypothesis*. We can think of L as being the cost of misclassifying a particular observation, which may vary. Note that, since we can add arbitrary constants to L and normalize so that it sums to 1 over all p, it is equivalent to assume the algorithm can solve

$$\max_{h\in\mathcal{H}} \sum_{p} \mathbf{1}[h(p) = y(p)]D(p), \tag{4.6}$$

for a probability distribution D over p. This provides an alternative interpretation, regarding the classifier seeking to make the correct guess with the highest possible probability.

We treat the process of finding the best fitting hypothesis as a black box. The purpose of this paper, however, is to understand how the algorithm designer might utilize from additional capabilities, and across a variety of environments. One question is which kinds of additional capabilities are necessary. The main ones we will discuss are:

- Constructing labels based on observations,
- Creating classifiers derived from solutions to the above maximization,
- Changing observations of p_t to $\hat{p}_t$.

One hypothesis class is of particular interest. Let H be a hyperplane in $\mathbb{R}^n$: $\exists\lambda \in \mathbb{R}^n$ and $\omega \in \mathbb{R}$ such that

$$\mathsf{H} = \left\{p \in \mathbb{R}^n \mid \lambda p = \omega\right\}.$$

Define H_+ as the close half space above H:

$$\mathsf{H}_+ = \left\{p \in \mathbb{R}^n \mid \lambda p \geq \omega\right\}.$$

Definition 4.2. *A single threshold (linear) classifier is a mapping*

$$h : \mathcal{P} \to A$$

where $\exists a_+, a_- \in A$ *such that*

$$h(p) = \begin{cases} a_+ & \text{if } \ p \in \mathsf{H}_+ \\ a_- & \text{if } \ p \notin \mathsf{H}_+. \end{cases}$$

We can define a classifier with multiple thresholds, built on multiple hyperplanes. As the number of hyperplanes increases, the classifier can assign different values of A over a finer partition. We use the number of hyperplanes associated with a classifier as the measure of complexity.

Definition 4.3. *If* $\tilde{\Gamma}$ *which is a collection of classifiers with at most* $\overline{N}$ *hyperplanes, we write* $\mathsf{cpx}(\tilde{\Gamma}) = \overline{N}$. *If* $\tilde{\Gamma}$ *admits a classifier with unlimited number of hyperplanes, then* $\mathsf{cpx}(\tilde{\Gamma}) = \infty$.

As we increase the number of threshold, the classifier can approximate any measurable function from $\mathcal{P}$ to A (Hornik, Stinchcombe, and White (1989)). If we allow $\tilde{\Gamma}$ to be any classifier over $\mathcal{P}$, then we essentially assume that the agent has an unlimited capability in sorting out $p \in \mathcal{P}$.

Definition 4.4. *Let Γ be the set of all classifiers, and $\tilde{\Gamma} \subset \Gamma$ denote a subset of classifiers. A statistical procedure or algorithm is an onto function*

$$\tau : \mathcal{D} \to \tilde{\Gamma},$$

where $\mathcal{D}$ is a set of histories, which consists of the realized p and $u(\theta, p, a)$, for all a $\mathcal{T}$ is the set of feasible algorithms.

An important substantive assumption here is that the algorithm can observe the ex-post utility for each action. This eliminates the need to consider experimentation incentives which would emerge if the designer would need to incur to obtain information about each a individually, leading to a bandit setting and adding significant complications. With a large number of agents, however, one could imagine giving a recommendation to a small fraction simply for the purpose of learning this value.[6] We wish to avoid this complication to maintain focus on the prediction problem.

One specification of $\mathcal{T}$ emerges from not having any restrictions on $\tilde{\Gamma}$ at all. In general, the set $\tilde{\Gamma}$ will be implicit in the description of the algorithm. Our main interest is in understanding which kinds of $\mathcal{T}$ allow for the buyer to approximate rational label y^R.

4.2. **Timing and Objectives.** An algorithm game takes the interaction in the stage game as a starting point, and considers the outcome when, instead of having the buyer's strategy emerging from Bayesian rationality, it instead emerges from fitting a model to past observations.

An algorithm game is a simultaneous move game under asymmetric information between the (rational) principal and the boundedly rational agent, built on the "expanded" stage game.

- A_{-1}. Nature first selects the parameters κ of the underlying game from $\mathcal{K}$ according to a prior distribution with a full support over $\mathcal{K}$, and informs only the principal.
- A_0. Conditioned on realized κ, the principal commits to some strategy σ. The agent commits an algorithm $\tau \in \mathcal{T}$ without observing $\kappa \in \mathcal{K}$.
- A_1. The extended stage game is played, with the agent's strategy in each period t being $\tau(D_t)(p)$ (i.e., the action specified by the algorithm following principal action p at time t), with the algorithm adding the observation (which includes p and ex-post utility following each agent action) to the dataset at the end of each period.

These actions determine the realized payoffs by each player, as described in the previous section; the expression for the payoffs of the seller and the buyer are (3.3) and (3.4), respectively, when $\sigma_{b,t}(a|p)$ is given by $\tau(D_t)(p)[a]$.

[6]Practically, this kind of experimentation does seem to be used when platforms resort to A/B testing; platforms very well may recommend new sellers in order to obtain information about them.

We consider the objectives of the strategic player and the algorithmic player separately. The former is straightforward; given a sequence (θ_t, p_t, a_t), the principal's payoff is simply the long run average, namely $\lim_{T\to\infty} \frac{1}{T} v(\theta_t, p_t, a_t, \kappa)$, maximizing this given κ. The algorithm designer also seeks to maximize long-run agent welfare. However, our aim is to capture a case where the algorithm designer seeks to use as little knowledge of the particular environment as possible, and therefore to apply across a variety of different settings. For instance, a retail platform may consist of a large number of different products, in which case the goal of the designer would be to ensure that the buyers do well in all of them. As such, whereas the ex-post payoff of the algorithm designer is again the long run average $\lim_{T\to\infty} \frac{1}{T} u(\theta_t, p_t, a_t, \kappa)$, the designer seeks to maximize this independently of κ.

As we will see, this objective for the algorithm designer justifies the algorithms' goal to seek to achieve the rational benchmark, and not to attempt to use first mover advantage to outperform it. We note that the comparison is potentially unfair because algorithms are more constrained in the decision rules that can be used. We therefore introduce a notion of rationality reflecting these limits:

Definition 4.5. *An algorithm τ is **constrained rational at** κ, if $\forall \epsilon, \delta > 0$, $\forall \sigma$, $\exists T$ such that $\forall t \geq T$,*

$$\mathbf{P}\left(v(\theta, p, \tau(D_t)(p))\sigma(p|\theta)\pi(\theta) \geq \max_{h\in\tilde{\Gamma}} v(\theta, p, h(p))\sigma(p|\theta)\pi(\theta) - \epsilon \right) \geq 1 - \delta.$$

*An algorithm τ is **fully rational** if $\tilde{\Gamma}$ is replaced by the set of all $h : \mathcal{P} \to A$.*

The "constrained" qualifier is due to the limits on the strategies that can be chosen by the agent. A fully rational agent would act optimally; a constrained rational algorithm yields actions are as optimally as possible, given that its output must be within the expanded model class $\tilde{\Gamma}$.

We often regard $\gamma \in \tilde{\Gamma}$ as a forecasting rule and τ as a formal procedure to construct a (strong) forecasting rule. If τ emulates the rational behavior, then the agent behaves as if he has perfect foresight about the seller's strategy σ in the long run, and chooses an optimal response from the set of feasible forecasting rule $\tilde{\Gamma}$.

In order to learn the equililbrium outcome $y^R(\sigma^R, \cdot)$, σ^R must be a best response to the decision rule induced by the algorithm in the long run.

Definition 4.6. *An outcome $(\overline{\sigma}, \tau)$ of the algorithm game **emulates** (σ^R, y^R) **of the underlying stage game,** if $\overline{\sigma} = \sigma^R$ and τ is fully rational.*

The substance of the definition is that σ^R is a best response to τ for almost all $\kappa \in K$. Then, along the equilibrium path of the algorithm game, the agent behaves as if he perfectly foresees σ^R and responds optimally subject to the feasibility constraint imposed by $\tilde{\Gamma}$.

5. Exploiting and Being Exploited

This section presents some preliminary observations which motivate our exercise subsequently. Specifically, we show first, that the set $\mathcal{H}$ should be taken to include at least single-threshold classifiers in order to outperform trivial decision rules which always give

the same recommendation. Second, we describe how seeking to provide good predictions across a variety of environments singles out the rational benchmark as the target for the algorithm, and not to seek to do better or worse than this. For the latter point, recall that the algorithm designer has first-mover advantage, since they are able to dictate how the buyer will respond to the seller's offers. Uncertainty over κ can mitigate the benefits of this advantage. In subsequent sections, we show how fitting a rich set of single-threshold classifiers, together with an apporpriate algorithm, can be used to return to the rational benchmark.

Note that the algorithm designer can always prescribe that some fixed classifier h be used, independently of the seller's actions. Not only is this the case, but in fact this would be the optimal strategy in many cases if κ could be conditioned on. More specifically, suppose $u(\theta, p, a, \kappa)$ is *independent* of θ, and weakly increasing (coordinatewise) in p, for each a (the latter of which would hold if, for instance, p were a menu of prices. Suppose further that this function is quasiconcave in p. The following simple result shows that in this case, at least (increasing) single threshold classifiers should be included:

Proposition 5.1. *Consider some κ where $u(\theta, p, 1, \kappa) - u(\theta, p, 0, \kappa)$ is constant in θ and weakly concave in p. Then if $u(\theta, p^*, 1, \kappa) > u(\theta, p^*, 0, \kappa)$, then there exists a single threshold classifier ensuring the strategic player chooses p^* with probability 1.*

The proof is straightforward, and follows immediately from an observation that the set of p at which the consumer chooses $a = 1$ is convex under the conditions of the proposition.

In order for the algorithm designer to improve upon a degenerate prescription to always choose $a = 0$, Proposition 5.1 suggests including at least single threshold classifiers which are increasing. Against the highlighted κ, such prescriptions would give the agent even higher commitment power than the rational benchmark. In order to maximize payoff against richer and richer κ, more and more classifiers should therefore be included to $\mathcal{H}$.

This raises the question of whether adding in these classifiers goes "too far." Namely, in seeking to maximize payoff against a rich set of possible κ, does this risk doing *worse* against others? In fact, it may be that the agent does *worse* than the rational benchmark.

Proposition 5.2. *Suppose that $\tilde{\Gamma}$ is the set of all single threshold classifiers, and suppose $\Theta = \{\theta_L, \theta_H\}$. Consider any κ satisfying the following:*

- *The principal's optimal $p - a$ pair when $\theta = \theta_L$ is $(p_L^*, 1)$*
- *The principal's optimal $p - a$ pair when $\theta = \theta_H$ is $(p_H^*, 0)$, with $p_L^* < p_H^*$.*
- *$v(\theta_H, p_H^*, 1, \kappa) = v(\theta_H, p_H^*, 0, \kappa)$,*
- *$v(\theta_L, p_L^*, 1, \kappa) \geq v(\theta_L, p_L^*, 0, \kappa)$, and*
- *$v(\theta, p, 1, \kappa) - v(\theta, p, 0, \kappa)$ increasing in p, for all θ.*

Then a policy arbitrarily close to the principal's optimal $p - a$ pair is implementable, even if this differs from the rational outcome under σ^R.

A setting where this principal-optimal action strategy differs from the rational outcome was first studied, to the best of our knowledge, in Rubinstein (1993). His setting satisfies the conditions of the proposition. Our proof adapts his arguments to the current setting (i.e., incorporating the statistical aspect of our exercise and beyond the application he considered, described in Section 3.3.1). We briefly describe the construction and the intuition. The reason the seller can profitably deviate in the previous proof is because

the new σ induces a *non-monotone* response from the agent optimally, even though this is not prescribed by σ^R. In contrast, decision rules with single-threshold classifiers must be monotone.[7]

In other words, we show that the principal can construct a strategy which ensures that the agent's utility as a function of p violates single-crossing. Now, if the seller were using the particular σ^* from the previous proof, then the rational response can be achieved via a double-threshold classifier, since there are only three optimal principal choices. But on the other hand, if the buyer were restricted to using single- or double-threshold classifiers, then one could find another strategy whereby the optimal response would be to use a triple-threshold classifier, via a similar scheme. If an agent is endowed with $\tilde{\Gamma}$ with $\mathsf{cpx}(\tilde{\Gamma}) < \infty$, there is no Nash equilibrium in the algorithm game that emulates (σ^R, y^R), because the principal can always exploit the limited perception of the pricing rules.

Briefly, we point out that argument also illustrates why the algorithm designer would also seek to include decreasing hyperplane classifiers, as well as increasing ones. In particular, if $v(\theta_L, p_L^*, 1, \kappa) > v(\theta_L, p_L^*, 0, \kappa)$ in the previous proposition, the proof shows that in fact the optimal rule to follow is a *decreasing* one, and so these should be included in order to outperform the $a = 0$ default.[8] With this in mind, we introduce the following richness condition:

Definition 5.3. *A set of parameters $\mathcal{K}$ satisfies richness if, for every distinct single-threshold classifier $h \in \mathcal{H}$, there exists a $\kappa_h \in \mathcal{K}$ such that h outperforms a trivial classifier given κ_h.*

The richness assumption motivates including at least the set of single threshold classifiers, as this is necessary to outperform degenerate predictions. The previous propositions suggest that this condition is not as restrictive as it may appear, as many natural specifications would feature this. On the other hand, Proposition 5.2 suggests that including these classifiers may be counterproductive, and lead to an even worse outcome than rational behavior.

6. Statement of the Main Result

We have already seen that some extension of $\mathcal{H}$ will in general be necessary in order to ensure that the algorithm induces the rational reply, as a principal can potentially exploit a departure from full rationality to increase payoffs. In addition, ensuring rationality would be trivial if the algorithm could use the parameter κ of the underlying game. In this case, $\mathcal{H}$ need only contain a single element, the rational reply to (σ, p). Due to the richness of the set of possible environments, $\mathcal{H}$ must be correspondingly rich; which, as we have seen, may provide scope for exploitation. We now consider the case where richer algorithms can be designed to counteract this.

We restrict the agent's classifier as emerging as the outcome of an *ensemble algorithm.*

[7]Even though $y^R(\sigma^R, p)$ is an element of $\tilde{\Gamma}$, $y^R(\sigma^*, p)$ is not an element of $\tilde{\Gamma}$. Since σ^* is the choice variable of the principal, the principal generates misspecification endogenously.

[8]In the case of a rational fringe as in Rubinstein (1993), this strict preference might emerge when there is heterogeneity in the value of the product when $\theta = L$.

Definition 6.1. *Classifier H is an ensemble of $\mathcal{H}$ if $\exists h_1, \ldots, h_K \in \mathcal{H}$ and $\alpha_1, \ldots, \alpha_K \geq 0$ such that*

$$H(\sigma, p) = \arg\max_a \sum_{k=1}^{K} \alpha_k \mathbf{1}[a = h_k(\sigma, p)]$$

Without loss of generality, we can assume that $\sum_{k=1}^{K} \alpha_k = 1$, since if not we can simply divide by this sum and obtain the same classifier.

We can interpret H as a weighted majority vote of $h_1, \ldots, h_K$. An ensemble algorithm constructs a classifier through a linear combination classifiers from $\mathcal{H}$. Since the final classifier is constructed through a basic arithmetic operation, one can easily construct an elaborate classifier from rudimentary classifiers. Ensemble algorithm has been remarkably successful in real world applications (Dietterich (2000)).

The algorithms produce an output ensemble classifier according to the following scheme:

- First, the loss function in (4.5), say L_1, or probability distribution in (4.6), say d_1, is taken to treat all observed principal actions symmetrically—that is, $L_1(p) = d_1(p) = 1/m$.
- At each stage $k = 1, \ldots$, the best fitting hypothesis is found by solving either (4.5) or (4.6). The best fitting hypothesis is referred to as h_k.
- The term α_k is then determined, possibly as a function of the objective of the best fitting hypothesis.
- Depending on h_k and α_k, the loss function L_k is updated to L_{k+1} (or, in the case of distributions, d_k is updated to d_{k+1}).
- After repeating this iteration K times, a classifier of the form of Definition 6.1 is output, which is used to determine the final choice of the agent.

The ability to use an ensemble algorithm allows additional richness in the set of classifiers that can be used. There remain, however, a number of challenges:

- Clearly, repeatedly solving the same problem will not yield different outcomes, and so to meaningfully expand $\mathcal{H}$ one needs to determine how to change the objective to be fit as well, and
- Weights must be specified in advance.

Both of these are *on top of* the need to potentially alter the observed p_t and determining the labels $y_t(\sigma, p_t)$ to use for the observations, since the observed utility-maximizing decision need not coincide with the rational one ex-post.

The main result of the paper is the following:

Theorem 6.2. *Let $\mathcal{H}$ consist of single-threshold classifiers, and suppose $|\mathcal{P}| < \infty$. Then there exists an algorithm such that the myopic agents play the rational best replies to the seller's choice, for all κ.*

The following corollary shows that this provides the optimal reply across $\kappa \in \mathcal{K}$ provided this set is rich:

Corollary 6.3. *Suppose $\mathcal{K}$ satisfies the richness condition (Definition 5.3) and that the algorithm designer seeks to ensure there is no κ such that the myopic agent does worse than the rational best reply (in the long run). Then the outcome of the interaction is the principal preferred equilibrium.*

The corollary provides a new justification for equilibrium selection in a number of settings, most notably informed principal settings which have long been known to be subject to multiplicity issues.

An important question is how to relax the assumption of $|\mathcal{P}| < \infty$. While clearly important, it is somewhat orthogonal to the other issues, and discussed in Section 7.4. This is done by introducing *approximations* which essentially make the observation space finite, and with the approximation becoming exact as the amount of data becomes infinite:

Definition 6.4. *Let $\{\mathcal{P}_{t,\lambda} : P^t \times A^t \to \Delta(P^t \times A^t)\}_{\lambda,t}$ be a collection of mappings, where $\lambda \in [0,1]$. Denote the image as $\overline{\mathcal{P}}_\lambda$. We call this an approximation given σ if, for all t and all $p^t \in P^t$, $\mathcal{P}_{t,\lambda}(p^t, a^t) \to (p^t, a^t)$ uniformly as $\lambda \to 0$, for all t, and in addition $\overline{\mathcal{P}}_\lambda < \infty$ for all λ, t. We call this an approximation if it is an approximation given σ, for all σ.*

Our result is to show that one can find a recursive ensemble algorithm for which the outcome of the game approximates rationality. More precisely, we construct algorithm $\tau^\lambda_{\hat{A}}$, which is parameterized by $\lambda > 0$. Given that our interest in the long run average payoff, as $T \to \infty$ the principal follows the equilibrium strategy even though the algorithmic agent may be boundedly rational.[9] In other words, the principal treats the agent as-if rational:

Proposition 6.5. *Suppose $\mathcal{H}$ is weakly learnable, and $y^R(\sigma, p)$ is the strict best response for each p in the support of $\sigma \in \Sigma$. There exists $\tau_{\hat{A}} \in \mathcal{T}$ such that the best response of the seller to $\tau_{\hat{A}}$ is σ^R. That is, there exists T and $\rho > 0$ such that*

$$\mathbf{P}\left(\tau_{\hat{A}}(D_t)(p) = y^R(\sigma^R, p) \ \forall t \geq T\right) \geq 1 - e^{-t\rho}$$

7. Constructing the Algorithm in the Main Result

This section describes the proof of the results in the previous section. The main condition necessary to ensure that this algorithm is well-defined and will converge to the best replies is *weak learnability*. We define this term below, and illustrate that it holds in our setting. The rest of the arguments relate to specifying the dataset for the algorithm, namely the labels used and the discretization referenced in Proposition 6.5, as well as describing the choice of weights and the objective in the "best-fitting hypothesis" problem described previously.

7.1. Weak Learnability.

The sufficient condition which ensures we can approximate an arbitrary decision rule using single-thresholds is *weak learnability*. Roughly speaking, weak learnability says that the hypothesis class can outperform someone who had some very minimal knowledge of the truth of the hypothesis. That is, it must be that the hypothesis class can do better than a someone who made a random guess, which would be made correct with some arbitrarily small probability. While this may seem minor—and indeed, it is certainly less stringent than requiring it can approximate the truth with high probability—the difficulty in achieving it is the fact that this guarantee must be uniform over all possible distributions.

We formally define this as follows:

[9]Prior versions of the paper showed that given any level of discretization, a sufficiently high discount factor ensures that the principal does not gain from exploiting the agent's bounded rationality.

Definition 7.1. *If $\overline{h}$ solves*

$$\sum_{p\in P(\sigma)} d(p)\mathbf{1}[y(p)=\overline{h}(p)] \geq \sum_{p\in P(\sigma)} d(p)\mathbf{1}[y(p)=h(p)] \qquad \forall h\in\mathcal{H},$$

$\overline{h}$ is an optimal weak hypothesis.

Definition 7.2. *A hypothesis class $\mathcal{H}$ is weakly learnable if, for every distribution f over observations $x\in X$ and labels $y(x)$, the optimal weak hypothesis satisfies:*

$$\sum_{x\in X} \mathbf{1}[\overline{h}(x)\neq y(x)]f(x) \leq \sum_{x\in X} \mathbf{E}_{\tilde{y}\sim B}[\mathbf{1}[\tilde{y}\neq y(x)]-\rho\mathbf{1}[\tilde{y}\neq y(x)]]f(x),$$

for some $\rho>0$ and some distribution B over A.

This condition reflects the idea that the classifier randomly guesses the label according to some distribution B, but is "flipped to being correct" with probability ρ. The right hand side describes the expected error in such a case, and the left hand side describes the error from the optimal weak classifier.

Weak learnability is tight, in the sense that if it fails, then *no* recursive ensemble algorithm can be built to approximate $y(x)$ based on $\mathcal{H}$ alone.[10] Perhaps more surprising is that it is tight, a fact which we discuss further in Section 7.3. For now, we simply mention that if we take $\mathcal{H}$, the set of single threshold classifiers is weakly learnable.

Proposition 7.3. *The set of single-threshold classifiers satisfies the weak learnability condition.*

Our proof uses the following important fact, which we prove in the Appendix: Any hypothesis class that *contains all label permutations* can at least match the random guess guarantee. The proof of this intermediate lemma uses a duality argument in order to show that no distribution can lead to a lower payoff when this condition is satisfied. Importantly, however, this is true for *any* hypothesis class, including the trivial one. This observation allows us to show that the added richness of single-threshold classifiers is sufficient to provide the additional gain over random guessing.

7.2. **Determining Labels.** The previous algorithm would be sufficient to construct a rational response if there were no uncertainty in the environment. However, if in fact randomness emerges, then the ex-post observation is not the rational reply, but in fact an estimated object. This section describes how these can be inferred from sample averages. For simplicity, we present the $|A|=2$ case, leaving the general case to the appendix.

We drop the assumption that the rational label is observed to construct another "intermediate" algorithm $\tau_{\hat{A}}$ before constructing the algorithm for Proposition 6.5. Now, the algorithm cannot observe σ, but observes the realized sign $\hat{y}_t(p)$ of

$$\sum_{t'=1}^{t-1} v(\theta,p,1)-v(\theta,p,0). \tag{7.7}$$

[10]For example, imagine $\mathcal{H}$ only consists of trivial classifiers. A corollary of B.1 is that these classifiers can do as well as a random guesser. However, it is clear that they cannot do better, as they are restricted to giving the same guess to all possible p, whereas this need not be true for a random guesser who is correct with an added probability ρ.

That is,

$$\hat{y}_t(p) = \begin{cases} 1 & \text{if } \sum_{t'=1}^{t-1} v(\theta, p, 1) - v(\theta, p, 0). \geq 0 \\ -1 & \text{if } \sum_{t'=1}^{t-1} v(\theta, p, 1) - v(\theta, p, 0). < 0. \end{cases}$$

Let $f_t^y(p)$ be the empirical probability that $\hat{y}_t(p) = 1$ at the beginning of period t. Thus, $\hat{y}_t(p) = -1$ with probability $1 - f_t^y(p)$. Given $\{d_t(p), \hat{y}_t(p)\}_p$, h_t solves

$$\max_{h \in \mathcal{H}} \sum_p h(p) d_t(p) [1 \cdot f_t^y(p) - 1 \cdot (1 - f_t^y(p))]$$

and

$$\hat{\epsilon}_t = \sum_p d_t(p) \left[f_t^y(p) \mathbb{I}(h(p) = 1) + (1 - f_t^y(p)) \mathbb{I}(h(p) = -1) \right].$$

Using weak learnability, we can show that $\exists \rho > 0$ such that

$$\hat{\epsilon}_t \leq \frac{1}{2} - \rho.$$

Since $\hat{y}_t(p)$ has the full support over $\{-1, 1\}$ $\forall t \geq 1$,

$$\hat{\epsilon}_t > 0.$$

Given an algorithm τ_A with observed labels, we can therefore replace it with $\tau_{\hat{A}}$ which involves inferring the labels y, setting them equal to $\hat{y}_t$, for all $t \geq 1$.

Proposition 7.4. *Fix $\sigma \in \Sigma^G \subset \Sigma$ where $y^R(\sigma, p)$ is a strict best response at p. Then, $\forall \epsilon > 0$, $\exists T$ and $\exists \overline{\rho} > 0$ such that*

$$\mathbf{P}\left(\exists t \geq T,\ \tau_{\hat{A}}(D_t)(p) \neq y^R(\sigma, p) \right) \leq e^{-\overline{\rho} t}.$$

Proof. See Appendix. □

7.2.1. *Remarks.* The ordinal information (7.7) about the average quality is necessary. Without access to (7.7), the algorithm cannot estimate $y(\sigma, p)$, which is critical for emulating the rational behavior.

The information contained in (7.7) is coarse, because the algorithm does not take any cardinal information about the parameters of the underlying game. Without the cardinal information, the agent cannot implement the equilibrium strategy of the baseline game, which is a single threshold rule. Because the algorithm does not rely on parameter values of the underlying game, the algorithm is robust against specific details of the game, if the algorithm can function as intended by the decision maker.

7.3. **Convergence.** So far, we have shown that we can construct labels from data, and that for the hypothesis class of interest the weak learnability condition is satisfied. The last step to show the algorithm works, in the case where the set of possible p has finite support, is that the output of the algorithm will indeed converge to the rational reply, as dictated by the labels, provided the weights are specified correctly.

For this, we use the Adaptive Boosting algorithm, as introduced by Schapire and Freund (2012), to specify the α_k weights and the updates. The original Adaptive Boosting

algorithm only applies to the case of $|A| = 2$. To handle the case of $|A| > 2$, we appeal to a generalization introduced by Mukherjee and Schapire (2013).[11]

The arguments for these proofs follow from results in the machine learning literature (see Schapire and Freund (2012)), which we can apply to show that this algorithm can yield a response for which the misclassification probability vanishes.

Proposition 7.5. *Fix a positive integer $G < \infty$. $\exists \rho > 0$ such that $\forall \sigma \in \Sigma^G$ whose support is in P, $\mathbf{P}_{d_1}\left(H_k(p) \neq \hat{y}(p)\right) < e^{-\rho k}$.*

Proof. See Appendix E. □

The proof reveals that the rate at which the probability of misclassfication vanishes is determined entirely by the number of principal actions in the support of σ. Thus, the algorithm is efficient (in that it maintains an exponential rate of convergence).

7.4. **Discretization.** While $\tau_{\hat{A}}$ is designed to be robust against parametric details of the underlying problems, the algorithm is still vulnerable to strategic manipulation by the rational seller. The proof of Proposition 7.5 reveals that the rate of convergence is decreasing as the number of principal actions in the support of σ increases. The principal can randomize over a countably infinitely many number of actions to slow down the convergence rate, and take advantage of the slow rate. By the same token, $\tau_{\hat{A}}$ may not PAC learn uniformly the strategies of the seller. That said, such manipulation would be short lived, and therefore have limited gains.

We describe how to revise $\tau_{\hat{A}}$ accordingly to discretize the observation space. Instead of processing individual actions, we let $\tau_{\hat{A}}$ process a group of actions at a time, treating "close" actions as the same group. In principle, we want to partition $\mathcal{P}$ into a set of half-open rectangles intervals with size λ. More precisely, given some arbitrary λ, we can partition each dimension of a rectangle containing $\mathcal{P}$ into the collection of half open intervals of size $\lambda > 0$ with a possible exception of the last interval:

$$P_0^j = [\underline{p}, \underline{p} + \lambda), \ldots, P_{K_j^\lambda}^j = [\underline{p} + (K_j^\lambda - 1)\lambda), \overline{p}]$$

where K_j^λ is the number of elements in the partition and $j \in \{1, \ldots, n\}$ is a partiular dimension.

For each element in the partition, the algorithm receives an ordinal information about the average outcome from the decision, if it contains a principal action in the support of σ:

$$\hat{y}_t^\lambda(k) = a \text{ if } a = \arg\max \sum_{p \in P_k} v(\theta, p, a)$$

where p in the support of σ and P_k is the product of partition elements. Let $\tau_{\hat{A}}^\lambda$ be the algorithm obtained by replacing $\hat{y}_t(p)$ in $\tau_{\hat{A}}$ by $\hat{y}_t^\lambda(k)$. Note that as $\lambda \to 0$, the size of the individual elements in the partition shrinks and $\tau_{\hat{A}}^\lambda$ converges to $\tau_{\hat{A}}$ for a fixed σ.

[11]The $|A| > 2$ algorithm works for the $|A| = 2$ case, with one minor drawback, which is that the learnability constant must be computed in advance. While our work shows an algorithm exists, the computation of the learnability constant is more indirect and hence explicitly finding a parameter that works is more difficult.

Compared to τ_A and $\tau_{\hat{A}}$, $\tau^{\lambda}_{\hat{A}}$ takes only coarse information for two important reasons. First, the algorithm cannot differentiate two ps which are very close. This features makes the algorithm robust against strategic manipulation of the seller to slow down the speed of learning. Second, the algorithm cannot detect the precise consequence of its decision, but only the ordinal information of the past decision, aggregated over time. The second feature allows the alogorithm to operate with very little information about the details of the parameters of the underlying game.

7.4.1. *Smoothing.* Discretizing the action space as above is one way of ensuring that there are only a finite number of principal actions to worry about in the long run, and given a sufficiently fine discretization, any distinct p is distinguished by the algorithm. However, in principle, close principal actions may still be quite far in terms of payoffs, and only be distinguished in the long run. That is, there is no guarnatee that for a fixed horizon, that the algorithm is not grouping too many p possibilities. The issue is that the discretization approach uses no information about the agent's payoff function. Our other alternative describes more explicitly how *close* to rationality the buyer can achive, given some fixed discretization scheme.

The idea is the following: We add a small amount of noise to each observed p, with the amount of noise tending to 0 as the sample size grows large. Doing so allows us to show that the agent perceives the principal's strategy to have the property that $\mathbf{E}_\theta[u(a,\theta,p(a) \mid p]$ is uniformly equicontinuous in p. As a result, if the agent only seeks to use a strategy that is $\varepsilon-$optimal against σ, uniform equicontinuity implies that their best reply can essentially be collapsed within intervals.

It will additionally be important that the algorithm does not seek to make predictions at p values where the corresponding density would be estimated to be small. Hence a second step will be to determine whether a p realization occur in a region with sufficiently large probability, where the "sufficient" amount will also tend to 0 as the amount of data grows large.

Formally, suppose the platform observes data $((p_1,y),\dots,(p_n,y))$. Let $z_{\eta,i} \in B_1(0)$ be an independent random vector distributed according to:

$$\phi_\eta(z) = \frac{1}{K}\exp\left(-\frac{1}{1-\left|z/\eta\right|^2}\right)\frac{1}{\eta^{|A|-1}},$$

where K is a constant which ensures ϕ_η integrates to 1. Our first augmentation is the following:

- Replace the observed $p_1,\dots,p_n$ with $\hat{p}_1,\dots,\hat{p}_n$, where $\hat{p}_i = p_i + z_{\eta,i}$, with $z_{\eta,i}$ distributed according to the above.

Second, it turns out that the above smoothing operation only works if the density is sufficiently large. Otherwise, the smoothing noise has too much power.

- For any $\tilde{p} = (\tilde{p}_a)_{a\in A\setminus a_0}$ drawn, estimate the event that $\tilde{\sigma}_\eta(\tilde{p}) < \gamma$ by fixing some δ small and determining whether menu(s) p with $\max_{a\in A\setminus\{a_0\}} \tilde{p}_a - p_a < \delta$ occurs with frequency at least $(2\delta)^{|A|-1}\gamma$. Recommend action a_0 for any such p.

As $\delta \to 0$, the condition holds if the density is at least γ. Together with the previous, we can show that if the buyer instead observes noisy principal actions, the perceived principal's strategy is sufficiently well-behaved to maintain the appropriate convergence for the algorithm.

Proposition 7.6. *Suppose the seller is restricted to choosing distributions which are either discrete or continuous. Consider an algorithm which can ensure that an ε-rational label is PAC-learnable, for any arbitrary $\varepsilon > 0$ given a finite number of possible principal actions. Then there exists a smoothing operation which maintains PAC-learnability of ε-rationality, for every $\varepsilon > 0$.*

The idea of the proposition is to use the smoothing operation to show that the algorithm perceives that the principal uses a σ such that $\mathbf{E}[u(a, \theta, p(a)) \mid p]$ is uniformly equicontinuous. Given that we seek ε-optimality, uniform equicontinuity allows us to essentially discretize the menu space, transforming the environment into a much simpler one.

There are two important properties of the transformation which allows us to ensure this works. The first is that, defining $\tilde{\sigma}_\eta(\cdot \mid \theta)$ to be the perceived p distribution of $p_i + z_i$, we have:

$$D^\alpha \tilde{\sigma}_\eta(p \mid \theta) = \int_P D^\alpha \phi_\eta(p - \tilde{p}) \sigma(\tilde{p} \mid \theta) d\tilde{p},$$

so that $\tilde{\sigma}_\eta$ inherits the smoothness properties of ϕ_η. The second is that, on any compact subset of P, we have $\sigma_\eta(\cdot \mid \theta) \to \sigma(\cdot \mid \theta)$ uniformly. Now, in order to obtain uniform continuity as $\eta \to 0$, it will be important that we can simultaneously ensure that the seller's strategy does not involve dramatic movements in the conditional probability. For instance, suppose the seller were to use the following strategy:

$$\sigma(p \mid G) = p(\sin\left(\frac{1}{p}\right) + 1), \sigma(p \mid B) = p(\sin\left(\frac{1}{p} - \pi\right) + 1),$$

defined on an interval $[0, \overline{p}]$ such that both densities integrate to 1. Then $\mathbf{P}[\theta = G \mid p] = 1$ if $p = \frac{1}{(2k+1/2)\pi}$ for some $k \in \mathbf{N}$, and 0 if $p = \frac{1}{(2k+1/2)\pi}$, for some $k \in \mathbf{N}$. As $k \to \infty$ (so that $p \to 0$), this oscillates infinitely often.

We handle the problem this example poses by only making non-degenerate predictions if the probability of using such principal actions is sufficiently high. That is, we "ignore" p realizations which only occur with low probability according to an estimated density.[12] Seeking to estimate the probability that all principal actions are within δ of p in order to estimate the density is just one way of doing this step; for instance, one could estimate the CDF $\tilde{\sigma}_\eta(p)$, and use the estimated density to determine whether the observations should be thrown away. Ultimately, however, given the compact $\mathcal{P}$, we can minimize the probability that this is done by using sufficiently low thresholds. As a result, it has a vanishing impact on PAC-learnability, as well as the seller's expected profit.

[12]One may wonder why this trick works; for instance, we do not obtain the result when $\sigma(p \mid G) = \sin\left(\frac{1}{p}\right) + 1, \sigma(p \mid B) = \sin\left(\frac{1}{p} - \pi\right) + 1$. However, unlike the previous example, these will fail the continuity requirement on the seller's strategy space, which is needed in the proof.

8. Review of Examples

We have yet to show that for any best response σ of the principal to $\tau_{\hat{A}}$, $(\sigma, \tau_{\hat{A}})$ constitutes a Nash equilibrium of the algorithm game, which emulates the Nash equilibrium of the underlying game, $(\sigma^R, y^R(\sigma^R, p))$.

8.1. **PAC Learnable.**

Definition 8.1. *τ is PAC (Probably Almost Correct) learnable if $\forall \sigma \in \Sigma$, $\forall \epsilon > 0$, $\exists T$ such that*

$$\mathbf{P}\left(\tau(D_t)(p) \neq y^R(\sigma, p) \qquad \forall t \geq T\right) < \epsilon.$$

PAC learnability is a sufficient condition for a Nash equilibrium of the algorithm game.

Proposition 8.2. *If τ is PAC learnable, then (σ^R, τ) is a Nash equilibrium of the algorithm game which emulates $(\sigma^R, y^R(\sigma^R, p))$.*

Proof. If τ is PAC learnable, then the agent learns σ accurately in the long run. Thus, the long run average expected payoff of the principal is

$$\mathcal{U}(\sigma, \tau) = \mathbf{E}_\theta u(\theta, \sigma, y^R(\sigma, \sigma(\theta)))$$

By the definition,

$$\sigma^R = \arg\max \mathbf{E}_\theta u(\theta, \sigma, y^R(\sigma, \sigma(\theta))).$$

By PAC learnability,

$$\lim_{t \to \infty} \tau(D_t)(p) = y^R(\sigma^R, p)$$

almost surely. Thus, (σ^R, τ) constitutes a Nash equilibrium which emulates $(\sigma^R, y^R(\sigma^R, p))$. □

Proposition 7.4 falls short of proving $\tau_{\hat{A}}$ is PAC learnable, because we need $y^R(\sigma, p)$ to be a strict best response.

Corollary 8.3. *Suppose that $y^R(\sigma, p)$ is a strict best response $\forall \sigma$. Then, $\tau_{\hat{A}}$ is PAC learnable.*

8.2. **Monopoly Market.**

In the example of Rubinstein (1993), $y^R(\sigma, p)$ is not a strict best response, if

$$\mathbf{E}_\theta v(\theta, p, a) = 0 \qquad \forall a \in A = \{1, -1\} \tag{8.8}$$

so that the agent is indifferent between accepting and rejecting p.

Corollary 8.4. *In the example of Rubinstein (1993), $\tau_{\hat{A}}$ is not PAC learnable of y^R.*

Yet, $\tau_{\hat{A}}$ is an equilibrium algorithm, because a rational principal would not use any σ which assigns positive probability $p > v_L$ satisfying (8.8).

Lemma 8.5. *Fix σ which assigns $p > v_L$ with positive probability, satisfying*

$$\mathbf{E}_\theta v(\theta, p, 1) \geq 0. \tag{8.9}$$

Then, the ex ante expected profit of the principal against $\tau_{\hat{A}}$ from σ is strictly smaller than from σ':

$$\mathcal{U}(\sigma^R, \tau_{\hat{A}}) > \mathcal{U}(\sigma, \tau_{\hat{A}}).$$

Proof. See Appendix D. □

Proposition 8.6. *In the example of Rubinstein (1993), if σ is a best response to $\tau_{\hat{A}}$, then $(\sigma, \tau_{\hat{A}})$ is a Nash equilibrium of the algorithm game, which emulates $(\sigma^R, y^R(\sigma^R, p))$.*

Proof. Lemma 8.5 implies that again $\tau_{\hat{A}}$, the principal will not use σ which assigns a positive probability to p so that both 1 and -1 are best responses. Thus, if σ is a best response to $\tau_{\hat{A}}$, then $y^R(\sigma, p)$ is a strict best response $\forall p > v_L$. If $y^R(\sigma, p) = 1$ is a strict best response, then $\mathbf{E}v(\theta, \sigma(\theta), 1) > 0$. The argument in the proof of Lemma 8.5 implies that the principal can generate higher ex ante payoff by shifting the probability weight assigned to p to v_L when the true state is H. Thus, if σ is a best response, then

$$\mathbf{E}v(\theta, \sigma(\theta), 1) < 0$$

so that $y^R(\sigma, p) = -1$.

Combining the arguments, we conclude that if σ is a best response to $tau_{\hat{A}}$, then $\forall p > v_L$, $\sigma(p : L) = 0$, and the same logic implies that the principal should not trade if $\theta = H$.

Since $\forall p < v_L$, $\mathbf{E}[v : p] - p > 0$, $\tau_{\hat{A}}(D_t)(p) = 1$ and the principal generates negative payoff. Thus, if σ is a best response to $\tau_{\hat{A}}$, $\sigma(p : L) = 0$ $\forall p < v_L$.

Thus, if σ is a best response to $\tau_{\hat{A}}$, $\sigma(v_L : L) = 1$ and the principal choose a price at which no trading occurs, such as v_H, if the true state is H. The trading occurs only under $\theta = L$ and the delivery price is v_L, as in $(\sigma^R, y^R(\sigma^R, p))$. Thus, $(\sigma, \tau_{\hat{A}})$ emulates $(\sigma^R, y^R(\sigma^R, p))$. □

8.3. **Labor Market Signaling.** The firm's objective function is to forecast the productivity of the worker:

$$v(\theta, p, a) = -(\theta - a)^2$$

If A is a real line, then

$$y^R(\sigma, p) = \arg\max_{a \in A} \mathbf{E}_\theta \left[v(\theta, p, a) : p, \sigma \right]$$

where the posterior distribution over θ is calculated via Bayes rule from σ and the prior over θ. Strict concavity of v implies that $y^R(\sigma, p)$ is a strict best response $\forall \sigma, p$.

Without loss of generality, we consider a single threshold decision rule parameterized by (a^+, a^-, p^0):

$$h(p) = \begin{cases} a^+ & \text{if } \ p \geq p^0 \\ a^- & \text{if } \ p < p^0. \end{cases}$$

Let $\mathcal{H}$ be the set of all single threshold decision rules. In each round, h_t solves

$$\max_{h \in \mathcal{H}} \mathbf{E}_\theta \left[v(\theta, p, a) : p, \sigma \right]$$

if the data includes σ. We construct τ_A accordingly. If σ is not observable by the algorithm, we estimate the posterior distribution of σ conditioned on each p to construct $\tau_{\hat{A}}$. If the agent learns $y^R(\sigma, p)$ eventually $\forall \sigma, p$, then the principal's choice σ^R

$$\mathbf{E}[u(\theta, p, y^R(\sigma, p)) : \sigma, p] = \sum_\theta \sum_p u(\theta, p, y^R(\sigma, p)) \sigma(p : \theta) \pi(\theta).$$

If σ^R entails separation by the high productivity worker, then the Riley outcome is the solution, that generates the largest ex ante expected surplus for the principal among all

separating equilibria. In order to satisfy the incentive constraint among different types of the principal, the principal with $\theta = H$ incurs the signaling cost. If the signaling cost outweighs the benefit of separation, then σ^R is the pooling equilibrium where both types of the workers takes the minimal signal.

The analysis is based upon the assumption that $y^R(\sigma, p)$ is a strict best response $\forall \sigma, p$. As $|A| = J < \infty$, $y^R(\sigma, p)$ may not be a strict best response for some σ and p. Let us assume that

$$A = \{a_0 = 0, a_1, \ldots, a_J\}$$

and $a_i - a_{i-1} = \Delta > 0$ and $a_F = 1 + \sup \Theta > 0$. Although $y^R(\sigma, p)$ may not be a strict best response for some σ and p, the set of best responses contains at most 2 elements, which differ by $\Delta > 0$. Abusing notation, let $y^R(\sigma, p)$ be the set of best responses, if the agent has multiple best responses at p. Applying the convergence result, we have $\exists T$ such that

$$\mathbf{P}\left(\exists y \in y^R(\sigma, p), y = \tau_{\hat{A}}(D_t)(p) \qquad \forall t \geq T\right) < e^{-\rho t}.$$

For a sufficiently small $\Delta > 0$, σ^R is either a strategy close to the Riley outcome, or the pooling equilibrium where both types of the principals choose the smallest value of p.

8.4. **Informed Principal.** The decision problem of the agent is to identify each pair (x, q) of payment x and cost q as an acceptable constract ($a = 1$) or not ($a = -1$). Without loss of generality, we can assume that the agent uses the single threshold linear classifier induced by hyperplane

$$\mathsf{H}(\lambda_x, \lambda_q, \omega) = \{(x, q) : \lambda_x x + \lambda_q q = \omega\}$$

and

$$h(x, q) = \begin{cases} 1 & \text{if } \ (x, q) \in \mathsf{H}^+(\lambda_x, \lambda_q, \omega) \\ -1 & \text{otherwise.} \end{cases}$$

We can construct $\tau_{\hat{A}}$ by estimating $\mathbf{E}v(\theta, p, a)$ for each (p, a).

Lemma 8.7. *Suppose that σ asssigns a positive probability to (x, q) where*

$$\mathbf{E}(q - \theta x : (q, x)) = 0$$

for $x > 0$. Then σ is not a best response to $\tau_{\hat{A}}$.

Proof. Let (x, q) be some offer such that the agent is indifferent between accepting and rejecting, so that:

$$q - \mathbf{E}[\theta \mid (x, q)]x = 0$$

The principal's expected payoff is found by taking the expectation of $u(\theta, (x, q), a)$ over all realizations of x, q. By the law of iterated expectations, this occurs if and only if the principal's payoff is maximized following *each* realization of (x, q). We claim the principal is *not* indifferent between actions following any such (x, q). Indeed, letting $\mathbf{E}[\theta \mid (x, q)] = r$, indifference implies:

$$(1 - r)f(I - rx) + rf(I - L + (1 - r)x) = (1 - r)f(I) + rf(I - L).$$

Note that equality holds if $f(y) = y$. This implies that both lotteries, whether or not the principal accepts, have the same expected values. However, if f is concave, then since $I > I - rx > I - L + (1-r)x > I - L$, it must be that the left hand side is strictly greater than the right hand side.

It follows that if indifference holds, the principal strictly prefers the agent accept the offer by slightly reducing x. □

Following the same logic as in the previous example, we conclude that if σ is a best response to $\tau_{\hat{A}}$, then

$$\tau_{\hat{A}}(D_t)(q,x) = y^R(\sigma,(q,x))$$

with probability 1. A best σ to $\tau_{\hat{A}}$ emulates $(\sigma^R, y^R(\sigma^R, p))$.

9. Conclusion

9.1. **Discussion of the model.** Several of our modelling choices are described below:

9.1.1. *Computational cost.* Our assumption is that the algorithm must be designed *before* observing the underlying parameters of the game. As illustrated in Proposition 5.2 below, the principal will typically faces an incentive to add possible actions to the support of σ in this game. And if the support of σ is large, finding an optimal threshold is a complex task. Because σ is endogenous, the optimization problem is even more complicated.

One might wonder why the algorithm is not reoptimized every time σ is chosen and κ is realized. This modelling choice can be justified by the introduction of small costs to writing an algorithm. To see this, note that the equilibrium value of the agents against σ is endogenous. So unless we impose a restriction on the set of feasible strategies of the principal, the computational cost of calculating a best response for every σ overwhelms any potential gain from playing the game. At first glance, the prediction of a subgame perfect equilibrium does not appear to be robust against a small computational cost (see, for instance, Rubinstein (1986)).

We have in mind a situation where the algorithm designer pays a small fixed cost for a computational *code*. This implies that it is prohibitively costly for them to seek to reoptimize against each individual principal they may face. We then search for an algorithm which can calculate a best response on behalf of the agents. If such an algorithm exists, and if the algorithm is sufficiently simple, then the agent behaves "as if" he is rational so that the best response of the monopolistic principal is the subgame perfect equilibrium strategy. By writing a flexible algorithm, the algorithm designer is able to respond to more strategies without incurring the additional costs.

9.1.2. *Coarse information.* The input of the algorithm are the data of the outcome, not the parameters of the game—namely, the price and the ex-post optimal decision given the price. The data from the outcome is coarse in the sense that the algorithm can use only the ordinal information of the outcome. For example, if the consumer surplus is positive, the algorithm can use the information that the surplus is positive (or negative), but not the information about the size of surplus. Relying on coarse information, the algorithm can operate over a broad class of games and its performance is not affected by the details of the games. Such robustness is particularly sought after, if the buyer has to design the algorithm based upon coarse information about the underlying game.

9.1.3. *Endogenously misspecified models.* An interesting question in our context is whether the algorithmic buyer in our model is correctly specified or not. A conventional learning algorithm aims to find the best fit model in a fixed class of models. The learning algorithm searches for a threshold rule in the class of single threshold decision rules which maximizes his expected return. In a setting with a large rational fringe, for instance, the equilibrium strategy of algorithmic buyer in a baseline model is a single threshold rule and therefore, $\mathcal{H}$ is correctly specified in the sense of Esponda and Pouzo (2014). If the model is correctly specified, we obtain the convergence to the rational behavior under general conditions (e.g., Marcet and Sargent (1989)).

In our case, however, the principal strategically chooses her strategy to render the model of an algorithmic buyer misspecified. Given the set of single threshold decision rules, the seller uses a strategy which requires two thresholds to correctly label the decision. Misspecification of the agents' model is endogenously generated by the strategic choice of the principal, which prevents the algorithm from learning to respond rationally to the strategy of the seller. The agents need an algorithm that can identify the best fit model efficiently, while expanding the model class from the decision rules with a single threshold to those with multiple thresholds.

Despite a large literature on learning in economics, there has been little progress of the investigation on the evolution of model classes.[13] Exploiting the recent development in machine learning literature, we construct a learning algorithm over the model class, which allows the algorithm (acting on behalf of agents) to respond rationally to a broad class of strategies of the principal. In the end, the principal finds it optimal to play the equilibrium strategy as if they were facing a rational agent.

9.2. **Final Comments.** In this paper, we have demonstrated how "as-if rational" behavior may emerge when the algorithm that induces market behavior can utilize single-threshold classifiers, with behavior determined by the outcome of a recursive ensemble algorithm (i.e., AdaBoost). As first noted by Rubinstein (1993), this need not be the case when their behavior follows from the *optimally* chosen single-threshold classifier, and despite the complexity involved with determining this strategy based on data alone in a non-strategic setting. This paper has articulated the following tradeoff in the design of statistical algorithms to mimic rationality: on the one hand, simply fitting a single-threshold classifier to data will fall short of rational play and be exploited by a seller. On the other hand, it may not be clear why this is the end of the story. By adding the ability to fit classifiers repeatedly and combining them in particular ways, we show how the rational benchmark can be restored. In this paper, we have taken as a black box the ability to fit these classifiers. But given this, our algorithm articulates exactly how to put these fitted classifiers together in order to construct one which can mimic rationality arbitrarily well. Going forward, given how productive the machine learning literature has been in terms of designing algorithms for the purposes of classification, we hope that our work will inspire further analysis of how these algorithms behave in strategic settings. Along these lines, we suspect that Rubinstein (1993) (or similar models) may be a useful laboratory for furthering this agenda beyond the issues we have looked at here.

[13]Cho and Kasa (2015) considered a learning model with multiple but fixed model classes, which makes it difficult to examine a long run evolution of model classes.

Appendix A. Proofs for Section 5

Proof of Proposition 5.1. Concave differences implies that the set

$$K = \{p : u(\theta, p, 1, \kappa) \geq u(\theta, p, 0, \kappa)\}$$

is a convex set; if $u(\theta, p_i, 1, \kappa) - u(\theta, p_i, 0, \kappa) \geq 0$ for $i = 1, 2$, then the same conclusion holds for $\alpha p_1 + (1-\alpha)p_2$ for all $\alpha \in [0, 1]$. Therefore, given any p^* on the boundary, the supporting hyperplane theorem implies that we can find a linear hyperplane (λ, ω) tangent to this set at p^*.

Suppose the algorithm designer prescribes that the agent choose $a = 1$ at any menu p such that $\lambda \cdot p \leq \omega$ and $a = 0$ otherwise. Note that having the agent choose $a = 1$ therefore requires choosing p where the principal would rather the agent choose action $a = 0$, by definition of K. Therefore, the strategic player cannot do any better than choosing $\sigma(p \mid \theta)$ which is a point mass at p^*. □

Proof of Proposition 5.2. The ideas in this proof are largely borrowed from Rubinstein (1993), accommodating two additional features of our enviroment: (a) need to infer the strategy from observed data and (b) the generalized setting, but we provide the proof for completeness. We construct a strategy σ^* for the principal that generates higher payoff than the equilibrium strategy σ^R, thus deriving the contradiction that σ^R is a best response to τ in the long run. More precisely, define $(p^*_\theta, a(\theta))$ to be the principal payoff-maximizing strategy. We show that the principal can induce the agent to choose $a(\theta) \neq y_R(p^*_\theta)$.

Fix $\epsilon > 0$ small First suppose $v(\theta_L, p^*_L, a_1, \kappa) = v(\theta_L, p^*_L, a_0, \kappa)$. Let $\tilde{p} \in (p_L, p_H)$ satisfies $v(\theta_L, \tilde{p}, 1, \kappa) < v(\theta_L, \tilde{p}, 0, \kappa)$. (If p is multidimensional, we can take $\tilde{p}$ tp be on the line segment connecting p^*_L and p^*_H) Set $\eta = v(\theta_L, \tilde{p}, 0, \kappa) - v(\theta_L, \tilde{p}, 1, \kappa) > 0$. We then choose $\epsilon, \epsilon_H, \epsilon_L > 0$ to satisfy

$$\pi(H)\epsilon_H < \pi(L)\epsilon_L, \tag{A.10}$$

and such that

$$\frac{\epsilon_L}{\epsilon_L + \eta} < \epsilon < \frac{\pi(L)\epsilon_L - \pi(H)\epsilon_H}{\pi(L)\epsilon_L}. \tag{A.11}$$

Under the increasing differences assumption, we can find $p_i(\varepsilon_i)$ such that

$$\varepsilon_i = v(\theta_i, p_i(\varepsilon_i), 1, \kappa) - v(\theta_i, p_i(\varepsilon_i), 0, \kappa).$$

Consider the following randomized pricing rule σ^* of the seller: in state H, $\tilde{p}_H(\epsilon_H)$ is chosen with probability 1. In state L, $p_L(\epsilon_L)$ is chosen with probability $1 - \epsilon$ and $\tilde{p}$ with probability ϵ.

Under this strategy, the optimal response following $\tilde{p}$ is 0, and this does not vanish as all other parameters tend to 0. However, the ex-post optimal decisions are 1 for both $\tilde{p}_L(\epsilon_L)$ and $\tilde{p}_H(\epsilon_H)$. Nevertheless, (A.11) implies first, the decisionmaker prefers to choose $a = 1$ if and only if $\tilde{p}_L(\epsilon_L)$ than choose $a = 1$ if and only if $\tilde{p}_H(\epsilon_H)$; and second, that the loss from choosing $a = 1$ following $\tilde{p}$ is larger than the loss from choosing $a = 0$ at $\tilde{p}_L(\epsilon_L)$. Putting this together, and taking $\epsilon, \epsilon_L, \epsilon_H \to 0$ shows this policy approximates the principal's optimum, as desired.

The case of $v(\theta_L, p^*_L, a_1, \kappa) > v(\theta_L, p^*_L, a_0, \kappa)$ is even more straightforward, since in this case the gain from choosing a_1 is non-vanishing, meaning that we can set $\varepsilon_L = 0$.

The verification that the optimal rule converges to this threshold when emerging from data is straightforward; any recursive learning algorithm generates $\{\phi_t\}$ which converges to $\phi \in \left(v_L - \epsilon_L, \frac{v_H + v_L}{2}\right)$ to emulate the best response of type 1 buyer against σ. Thus, the long run average payoff against such algorithm should be bounded from below by $\mathcal{U}^*_p - \epsilon$. □

Appendix B. Weak Learnability Proofs

Lemma B.1. *Let $\mathcal{H}$ be an arbitrary hypothesis class with the property that for every $h \in \mathcal{H}$ and every permutation $\pi : A \to A$, the composition $\pi \circ h$ is contained in $\mathcal{H}$. Then this hypothesis class can do at least as well as a uniform random guesser.*

Proof. Let Π be the set of all possible permutations on A, noting that $|\Pi| = k!$. Fix an arbitrary classifier $h \in \mathcal{H}$, and define $h^\pi = \pi \circ h$. Let $c_{j,y}$ be the cost of assigning label y to price p_j. Define

$$\sum_{\pi \in \Pi} c_{j,h^\pi(p_j)} = \overline{c}_j.$$

In particular, note that this is invariant to the true label of j. As a result, the random guesser's expected payoff on observation j is is $\overline{c}_j/k!$. To see this, note that $h(p_j)$ gives some fixed guess regarding the label of price p_j. Then randomizing over permutations is equivalent to randomizing over labels, as there are an equal number of permutations which flip the label according to $h(p_j)$ and every other label.

We therefore obtain the following matrix equation, for an arbitrary $\rho \in (0, \infty)$, where the number of columns is $k!$ and the number of rows is the number of possible prices.

$$\left(\begin{array}{c|ccc|c} c_{j,h(p)} & & & & c_{j,h^\pi(p)} \\ -\overline{c}_j/k! & \cdots & & \cdots & -\overline{c}_j/k! \\ & & & & \end{array}\right) \cdot \begin{pmatrix} \frac{\rho}{k!} \\ \vdots \\ \frac{\rho}{k!} \end{pmatrix} = \mathbf{0}$$

Also note that:

$$(1/\rho, \cdots, 1/\rho) \cdot \begin{pmatrix} \frac{\rho}{k!} \\ \vdots \\ \frac{\rho}{k!} \end{pmatrix} = 1$$

So as long as $\rho > 0$, by the theorem of the alternative, we therefore cannot have that a vector $\mathbf{x}$ exists with:

$$\left(\begin{array}{ccc} c_{j,h(p)} - \overline{c}_j/k! & & \\ \hline \vdots & & \vdots \\ \vdots & & \vdots \\ \hline & & c_{j,h^\pi(p)} - \overline{c}_j/k! \end{array}\right) \cdot \mathbf{x} \geq \begin{pmatrix} \frac{1}{\rho} \\ \vdots \\ \frac{1}{\rho} \end{pmatrix}.$$

Let $D(p)$ be an arbitrary distribution. Since $\sum_{p \in P} D(p) = 1$, this implies we can find some π such that:

$$\left(\sum_{p_j \in P} D(p_j)(c_{j,h^\pi(p_j)} - \frac{\overline{c}_j}{k!})\right) < \frac{1}{\rho}.$$

Taking $\rho \to \infty$ and rearranging gives:

$$\left(\mathbf{E}_{p \sim D}[c_{j,h^\pi(p_j)}]\right) \leq \mathbf{E}_{j \sim D}\left[\frac{\overline{c}_j}{k!}\right]$$

Recalling again that the right hand side of this inequality is the payoff of the random guesser, we have shown that for every possible distribution over prices, we can find some permutation which delivers a cost bounded above by the random guesser. This proves the Lemma. □

Proof of Proposition 7.3. Let $\mathcal{H}$ to be the set of hyperplane classifiers,. We prove this by contradiction. If there were no universal lower bound on the error, then we would have, for all ρ, a distribution D_ρ and cost $c^\rho_{j,y}$ (without loss normalized to be on the unit sphere themselves) with the property that:

$$\max_{h \in \mathcal{H}} \sum_{p \in P} D_\rho(p) c^\rho_{j,h(p)} < U^\rho_c,$$

where U^ρ_c is the payoff of the uniform random guesser who is correct with added probability ρ. Taking $\rho \to 0$ and passing to a subsequence if necessary, compactness of the unit sphere implies that we can find a distribution D^* and cost function c^* such that:

$$\max_{h \in \mathcal{H}} \sum_{p \in P} D^*(p) c^*_{j,h(p)} = U^0_c,$$

where we note by Lemma B.1 that at least this bound can be obtained by permutation the labels if necessary. We will arrive at a contradiction by exhibiting a single-hyerplane classifier that achieves a strictly better accuracy, given D^*. Note that $\mathcal{H}$ contains the set of "trivial" classifiers, which give all menus the same label. Also note that the only non-trivial case to consider is when there are at least two prices in the support of D^*; if there were only one price, then simply choosing the prediction corresponding to the label on that price would yield a perfect fit. Since, by assumption, no classifier does better than random guessing, it must be the case in particular that each trivial classifier cannot exceed the random-guess bound. On the other hand, by our previous result, we know there *does* exist a trivial classifier which achieves at least this bound, for *any* D supported on P.

Let $P = \{p_1, \ldots, p_k\}$ be the set of prices supporting D^*, and let $\tilde{p} \in P$ be a price in that is also an extreme point of the convex hull of P. Without loss of generality, assume that $\tilde{p}$ is nontrivial, in the sense that it does not give the same cost to all labels. Note that indeed, this is without loss, since for any such price, the choice of classification is irrelevant.[14] Note that $\tilde{p}$ is not in the convex hull of $P\backslash\{\tilde{p}\}$. Therefore, by the separating hyperplane theorem, we can find an $h \in \mathcal{H}$ which (strictly) separates $\tilde{p}$ from $P\backslash\{\tilde{p}\}$. Denote such a hyperplane by h^*, and note that the set of hyperplane classifiers contains classifiers which assign *any* two labels (possibly the same label) to prices depending on which side of h^* they lie on.

Also note that, again by our previous result, a trivial classifier supported on $P\backslash\{\tilde{p}\}$ can achieve the random guess guaranatee if p is distributed according to the conditional distribution on this set. In other words, our prior lemma implies that there exists $y^* \in A$ such that:

$$\sum_{p_j \in P\backslash\tilde{p}} \frac{D^*(p_j)}{\sum_{q\in P\backslash\tilde{P}} D^*(q)} c^*_{j,y^*} = U^0_{c^*}.$$

On the other hand, a classifier which separates $p_{\tilde{\jmath}}$ from the other prices can fit $p_{\tilde{\jmath}}$ perfectly. Thus we must have

$$c_{\tilde{\jmath},y_{\tilde{\jmath}}} < \mathbf{E}_{\hat{y}\sim\mathrm{Unif}}[c_{\tilde{\jmath},\hat{y}}].$$

So consider the hyperplane classifier which predicts $\tilde{y}$ for $\tilde{p}$, and y^* for $p \in P\backslash\{\tilde{p}\}$, i.e., depending on which side of h^* they are on (acknowledging that this may be a trivial classifier). Denote the resulting classifier by h. For this single-hyperplane classifier, we have

$$\sum_{p_j\in P} D^*(p_j)c_{j,h(p_j)} = D^*(p_{\tilde{\jmath}})c_{\tilde{\jmath},y_{\tilde{\jmath}}} + \left(\sum_{q\in P\backslash\{p_{\tilde{\jmath}}\}} D^*(q)\right) \sum_{p_k\in P\backslash\{p_{\tilde{\jmath}}\}} \frac{D^*(p_k)}{\sum_{q\in P\backslash\{p_{\tilde{\jmath}}\}} D^*(q)} c_{k,y^*} > U^0_{c^*},$$

where the inequality holds since the single-threshold classifier does strictly better on some non-trivial price, and as well on all other prices. This completes the proof. □

Appendix C. Proofs for Determining Labels (Section 7.2)

Proof of Proposition 7.4. It suffices to show that $\forall\epsilon > 0$, $\forall\sigma \in \Sigma$, $\exists T(\sigma,\epsilon)$ such that

$$\mathbf{P}\left(H_t(p) \neq y^R(\sigma,p) \qquad \forall t \geq T(\sigma,\epsilon)\right) \leq \epsilon.$$

By the law of large numbers,

$$\hat{v}_t(p,a) \to v(p,a) = \mathbf{E}_\theta[v(\theta,p,a) : p,\sigma].$$

Under the assumption that $y^R(\sigma,p)$ is a strict best response,

$$\lim_{t\to\infty} \hat{y}_t(p) \to y^R(\sigma,p)$$

[14] If all prices are trivial, then we will achieve a contradiction, because that implies that the classifier does do at least as well as the edge-over-random guesser, since all classifiers achieve the same payoff.

almost surely. Invoking Cramér's theorem, we have that $\forall \epsilon > 0$, $\exists \rho(\epsilon, \sigma) > 0$ and $T(\epsilon, \sigma)$ such that

$$\mathbf{P}\left(\exists t \geq T(\epsilon, \sigma), \hat{y}_t(p) \neq y^R(\sigma, p)\right) \leq e^{-t\rho(\epsilon,\sigma)}.$$

The convergence rate function $\rho(\epsilon, \sigma)$ is determined by the empirical posterior distribution $\{\hat{v}_t(p, a)\}_{p \in \mathcal{P}(\sigma)}$. Since $\sigma \in \Sigma^G \subset \Sigma$ for a positive integer G, the empirical the multinomial probability distribution over θ.

Let $\hat{\pi}_t(\theta : p)$ be the empirical probability distribution over Θ following t rounds of observations. By the law of large numbers, $\hat{\pi}_t(\theta : p) \to \pi(\theta : p)$ computed via Bayes rule from the prior distribution over θ and σ. Write $\Theta = (\theta_1, \ldots, \theta_{|\Theta|})$. Given $\epsilon = (\epsilon, \ldots, \epsilon) \in \mathbb{R}^{|\Theta|}$, the rate function of the multinomial distribution is

$$\sum_{i=1}^{|\Theta|} \epsilon \log \frac{\epsilon}{p(\theta)}$$

where $p(\theta)$ is the probability that θ is realized. Since $\sum_\theta p(\theta) = 1$,

$$\sum_{i=1}^{|\Theta|} \epsilon \log \frac{\epsilon}{p(\theta)} \geq \prod_{i=1}^{|\Theta|} \epsilon \log \frac{\epsilon}{1/|\Theta|} = \prod_{i=1}^{|\Theta|} \epsilon \log \epsilon |\Theta| > 0.$$

Note that the right hand side is independent of σ. Thus, we can choose $\rho(\epsilon) \leq \rho(\epsilon, \sigma)$ uniformly over σ, which is strictly increasing with respect to $\epsilon > 0$. We choose $T(\epsilon)$ independently of σ as well.

Define an event

$$\mathcal{L} = \left\{\hat{y}_t(p) = y^R(\sigma, p) \qquad \forall t \geq T(\epsilon)\right\}$$

We know that

$$\mathbf{P}(\mathcal{L}) \geq 1 - e^{-t\rho(\epsilon)}.$$

Fix $t > T(\epsilon)$. We have

$$\begin{aligned}
& \mathbf{P}\left(\hat{H}_t(p) \neq y^R(\sigma, p)\right) \\
= \; & \mathbf{P}\left(\hat{H}_t(p) \neq y^R(\sigma, p) : \mathcal{L}\right)\mathbf{P}(\mathcal{L}) + \mathbf{P}\left(\hat{H}_t(p) \neq y^R(\sigma, p) : \mathcal{L}^c\right)\mathbf{P}(\mathcal{L}^c) \\
\leq \; & \mathbf{P}\left(\hat{H}_t(p) \neq y^R(\sigma, p) : \mathcal{L}\right) + \mathbf{P}(\mathcal{L}^c) \\
\leq \; & \mathbf{P}\left(\hat{H}_t(p) \neq y^R(\sigma, p) : \mathcal{L}\right) + e^{-t\rho(\epsilon)}.
\end{aligned}$$

Following the same logic as in the proof of Proposition 7.5, we can show that $\exists \gamma(G) > 0$ such that

$$\hat{Z}_t \leq 1 - \gamma(G) \qquad \forall t \geq 1 \tag{C.12}$$

under $\tau_{\hat{A}}$.

Recall that

$$F_a(p) = \sum_{s=1}^{t} \alpha_s \mathbf{1}(h_s(p) = a).$$

Similarly, we define

$$\hat{F}_a(p) = \sum_{s=1}^{t} \hat{\alpha}_s \mathbf{1}(h_s(p) = a).$$

Following the same logic as in the proof of Proposition 7.5, we know that if $\hat{H}_t(p) \neq y^R(p)$,

$$\hat{F}_{y^R(\sigma,p)}(p) + \sum_{a \neq y^R(\sigma,p)} \hat{F}_a(p) > 0.$$

Thus,

$$\begin{aligned}\mathbf{1}(\hat{H}_t(p) \neq y^R(\sigma,p)) &\leq \mathbf{1}\left(\hat{F}_{y^R(\sigma,p)}(p) + \sum_{a\neq y^R(\sigma,p)} \hat{F}_a(p)\right)\\ &\leq \exp\left(\hat{F}_{y^R(\sigma,p)}(p) + \sum_{a\neq y^R(\sigma,p)} \hat{F}_a(p)\right).\end{aligned}$$

Conditioned on event $\mathcal{L}$,

$$\hat{y}_t(p) = y^R(\sigma,p) \qquad \forall t \geq T(\epsilon).$$

We can write for $t \geq T(\epsilon)$,

$$\begin{aligned}d_{t+1}(p) &= \frac{\hat{d}_t(p)\exp(\alpha_t(\mathbf{1}(h_t(p)\neq \hat{y}_t(p)) - \mathbf{1}(h_t(p)=\hat{y}_t(p))))}{\hat{Z}_t}\\ &= \frac{\hat{d}_t(p)\exp(\alpha_t(\mathbf{1}(h_t(p)\neq y^R(\sigma,p)) - \mathbf{1}(h_t(p)=y^R(\sigma,p))))}{\hat{Z}_t}\\ &= \frac{d_{T(\epsilon)}(p)\exp(\sum_{s=T(\epsilon)}^t \alpha_s(\mathbf{1}(h_s(p)\neq y^R(\sigma,p)) - \mathbf{1}(h_s(p)=y^R(\sigma,p))))}{\prod_{s=T(\epsilon)}^t \hat{Z}_t}.\end{aligned}$$

Thus,

$$\begin{aligned}&\prod_{s=T(\epsilon)}^t \hat{Z}_t\\ =\ & \sum_p d_{T(\epsilon)}(p)\exp\left[\sum_{s=T(\epsilon)}^t \alpha_s(\mathbf{1}(h_s(p)\neq y^R(\sigma,p)) - \mathbf{1}(h_s(p)=y^R(\sigma,p)))\right]\\ \geq\ & \left(\min_{p\in\mathcal{P}(\sigma)} d_{T(\epsilon)}(p)\right)\sum_p \exp\left[\sum_{s=T(\epsilon)}^t \alpha_s(\mathbf{1}(h_s(p)\neq y^R(\sigma,p)) - \mathbf{1}(h_s(p)=y^R(\sigma,p)))\right].\end{aligned}$$

Since $d_1(p)$ is the uniform distribution over $\mathcal{P}(\sigma)$,

$$\min_{p\in\mathcal{P}(\sigma)} d_{T(\epsilon)}(p) > 0.$$

We can write

$$\begin{aligned}&\prod_{s=1}^t \hat{Z}_t = \prod_{s=T(\epsilon)}^t \hat{Z}_t \prod_{s=1}^{T(\epsilon)-1} \hat{Z}_t\\ \geq\ & \left(\min_{p\in\mathcal{P}(\sigma)} d_{T(\epsilon)}(p)\right)\sum_p \exp(\sum_{s=T(\epsilon)}^t \hat{\alpha}_s(\mathbf{1}(h_s(p)\neq y^R(\sigma,p)) - \mathbf{1}(h_s(p)=y^R(\sigma,p))))\prod_{s=1}^{T(\epsilon)-1}\hat{Z}_t\\ =\ & \frac{\left(\min_{p\in\mathcal{P}(\sigma)} d_{T(\epsilon)}(p)\right)\prod_{s=1}^{T(\epsilon)-1}\hat{Z}_t}{\sum_p \exp\left[\sum_{s=1}^{T(\epsilon)-1}\hat{\alpha}_s(\mathbf{1}(h_s(p)\neq y^R(\sigma,p)) - \mathbf{1}(h_s(p)=y^R(\sigma,p)))\right]}\\ &\qquad \times \sum_p \exp\left[\sum_{s=1}^t \hat{\alpha}_s(\mathbf{1}(h_s(p)\neq y^R(\sigma,p)) - \mathbf{1}(h_s(p)=y^R(\sigma,p)))\right]\end{aligned}$$

over $\mathcal{L}$. Define

$$M(\epsilon) = \frac{\left(\min_{p\in\mathcal{P}(\sigma)} d_{T(\epsilon)}(p)\right)\prod_{s=1}^{T(\epsilon)-1}\hat{Z}_t}{\sum_p \exp(\sum_{s=1}^{T(\epsilon)-1}\hat{\alpha}_s(\mathbf{1}(h_s(p)\neq y^R(\sigma,p)) - \mathbf{1}(h_s(p)=y^R(\sigma,p))))}$$

which is bounded away from 0.

Recall that

$$\begin{aligned} & \mathbf{P}(\hat{H}_t(p) \neq y^R(\sigma,p)) \\ \leq & \sum_p d_1(p) \exp(\sum_{s=1}^t \hat{\alpha}_s(\mathbf{1}(h_s(p) \neq y^R(\sigma,p)) - \mathbf{1}(h_s(p) = y^R(\sigma,p)))) \\ \leq & \frac{\prod_{s=1}^t \hat{Z}_t}{M(\epsilon)} \leq \frac{(1-\gamma(G))^t}{M(\epsilon)} \leq \frac{e^{-t\gamma(G)}}{M(\epsilon)}. \end{aligned}$$

Combining the probabilities over $\mathcal{L}$ and $\mathcal{L}^c$, we have that $\forall \epsilon$, $\forall \sigma \in \Sigma^G \subset \Sigma$, $\exists T(\epsilon)$, $\rho(\epsilon)$ and $\gamma(G)$ such that

$$\mathbf{P}\left(\exists t \geq T(\epsilon),\ \hat{H}_t(p) \neq y^R(\sigma,p)\right) \leq \frac{e^{-t\gamma(G)}}{M(\epsilon)} + e^{-t\rho(\epsilon)}.$$

We can choose $T > T(\epsilon)$ and $\overline{\rho}$ such that $\forall t \geq T$,

$$\frac{e^{-t\gamma(G)}}{M(\epsilon)} + e^{-t\rho(\epsilon)} \leq e^{-\overline{\rho}t}$$

which proves the proposition. □

Appendix D. Proofs for Examples

Proof of Lemma 8.5. It suffices to show that if $p > v_L$ and $\mathbf{E}(v|p) - p \geq 0$, then the expected profit from p is strictly less than $\pi_L v_L$. We write the proof in Rubinstein (1993) for the later reference. For any price p satisfying

$$\mathbf{P}(H|p)v_H + \mathbf{P}(L|p)v_L \geq p,$$

the revenue cannot exceed

$$\mathbf{P}(H|p)v_H + \mathbf{P}(L|p)v_L$$

but the cost is

$$\mathbf{P}(H|p)(1-r)c_2 + \mathbf{P}(H|p)rc_1.$$

Thus, the seller's expected profit is at most

$$\mathbf{P}(L|p)v_L + \mathbf{P}(H|p)((1-r)(v_H - c_2) + r(v_H - c_1))$$

Because of the lemon's problem,

$$(1-r)(v_H - c_2) + r(v_H - c_1) < 0$$

and

$$\mathbf{P}(H|p) > 0$$

to satisfy

$$\mathbf{P}(H|p)v_H + \mathbf{P}(L|p)v_L \geq p > v_L.$$

Integrating over p, we conclude that the ex ante profit is strictly less than $\pi_L v_L$.

□

Appendix E. Specifying the Algorithm Parameters and the Proof of Proposition 7.5.

E.1. **The $|A| = 2$ case.** The specification of the algorithm parameters coincides with the Adaptive Boosting algorithm of Schapire and Freund (2012). We first outline the parameters and then review, for completeness, the proof that we obtain the rate claimed int he proposition
At the kth stage (initializing with the uniform distribution), define

$$\epsilon_k = \mathbf{P}_{d_k}\left(h_k(p) \neq \hat{y}(p)\right) \tag{E.13}$$

as the probability that the optimal classifier h_k at k misclassifies p under d_k. If $\epsilon_k = 0$, then we stop the training and output h as the forecasting rule, which perfectly forecasts $y(\sigma,p)$.

Suppose that $\epsilon_k > 0$. Define[15]

$$\alpha_t = \frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t} \tag{E.14}$$

Define for each p in the support of σ, and each pair $(p, \hat{y}(p))$,

$$d_{k+1}(p) = \frac{d_k(p)\exp(-\alpha_k \hat{y}(p) h_k(p))}{Z_k}$$

where

$$Z_k = \sum_{p\in\overline{\mathcal{P}}_t} d_k(p)\exp(-\alpha_k\hat{y}(p)h_t(p)).$$

Given d_{k+1}, we can recursively define h_{k+1} and ϵ_{k+1}.

We now present the argument for convergence. Define

$$F_t(p) = \sum_{k=1}^{t}\alpha_k h_k(p).$$

Following the same recursive process described in Schapire and Freund (2012), we have

$$d_{t+1}(p) = \frac{d_1(p)\exp\left(-y(\sigma,p)\sum_{k=1}^t \alpha_k h_k(p)\right)}{\prod_{k=1}^t Z_k} = \frac{d_1(p)\exp(-y(\sigma,p)F_t(p))}{\prod_{k=1}^t Z_k}. \tag{E.15}$$

Following Schapire and Freund (2012), we can show that

$$\mathbf{P}\left(H_t(p)\neq y(\sigma,p)\right) = \mathbf{E}\sum_p d_1(p)\mathbb{I}(H_t(p)\neq y(\sigma,p)) \leq \mathbf{E}\sum_p d_1(p)\exp(-y(\sigma,p)F_t(p)),$$

and

$$\mathbf{P}(H_t(p)\neq y(\sigma,p)) = \mathbf{E}\prod_{k=1}^t Z_k.$$

Note

$$Z_k = \sum_p d_k(p)\exp\left(-y(\sigma,p)\alpha_k h_k(p)\right).$$

The rest of the proof follows from Schapire and Freund (2012), which we copy here for later reference.

$$\begin{aligned}
Z_t &= \sum_p d_t(p)\exp\left(-y(\sigma,p)\alpha_t h_t(p)\right)\\
&= \sum_{y(\sigma,p)h_t(p)=1} d_t(p)\exp\left(-\alpha_t\right) + \sum_{y(\sigma,p)h_t(p)=-1} d_t(p)\exp\left(-\alpha_t\right)\\
&= e^{-\alpha_t}(1-\epsilon_t) + e^{\alpha_t}\epsilon_t\\
&= e^{-\alpha_t}\left(\frac{1}{2}+\gamma_t\right) + e^{\alpha_t}\left(\frac{1}{2}-\gamma_t\right)\\
&= \sqrt{1-4\gamma_t^2}
\end{aligned}$$

where

$$\gamma_t = \frac{1}{2} - \epsilon_t.$$

By weak learnability, we know that γ_t is uniformly bounded away from 0: $\exists\gamma > 0$ such that

$$\gamma_t \geq \gamma \qquad \forall t \geq 1.$$

Recall that the maximum number of the elements in the support of σ is N. Thus,

$$d_{t+1}(p) = d_1(p)\prod_{k=1}^t\sqrt{1-4\gamma_t^2} \leq \frac{1}{N}\left(1-4\gamma^2\right)^{\frac{t}{2}} \leq \frac{1}{N}e^{-2\gamma^2 t}$$

[15]In general one should worry about this expression being defined; however, the fact that it is follows immediately from Weak Learnability.

where the right hand side converges to 0 at the exponential rate uniformly over p.

E.2. **The $|A| > 2$ case.** The specification of the algorithm can be found in Mukherjee and Schapire (2013). The proof provided here is somewhat more direct than their exposition, but borrows key ideas.

First, initialize $F_y^0(x_i) = 0$.

- From previous stage, take F_y^t.
- At stage t, find the $h \in \mathcal{H}$ solving:

$$\min_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}[h_t(x_i) = y_i]\left((e^{-\eta}-1)\sum_{\tilde{y}\neq y_i} e^{F_{\tilde{y}}^{t-1}-F_{y_i}^{t-1}}\right) + \mathbf{1}[h_t(x_i)\neq y_i](e^{\eta}-1)e^{\eta(F_{h_t(x_i)}^t - F_y^{t-1}(x_i))}.$$

- Define $F_y^t(x_i) = \sum_{s=1}^{t} \mathbf{1}[h_t(x_i) = y]$.

The final prediction is $H_t(x_i) = \arg\max_{\tilde{y}} \sum_{t=1}^{T} \mathbf{1}[h_t(x_i) = \tilde{y}]$.

The weak learnability condition says that the hypothesis class can outperform a random guesser that does better than some γ, where we allow for a potentially asymmetric cost of making different errors.

We now show convergence to the rational rule:

Step 1: Bounding The Mistakes: This step is as previous. We have

$$\sum_{i=1}^{m} \mathbf{1}[H_t(x_i)\neq y_i] \leq \sum_{i=1}^{m} e^{\sum_{\tilde{y}\neq y_i} F_{\tilde{y}_i}^T(x_i) - F_{y_i}^T(x_i)}.$$

Indeed, the exponential is positive, so this inequality holds when y_i is labelled correctly, and if the label is incorrect, then that means that some $\tilde{y}_i$ satisfies $F_{\tilde{y}_i}^T(x_i) > F_{y_i}^T(x_i)$. Since all $F_y^T(x_i)$ are positive, the exponent is positive if x_i is labelled incorrectly, meaning the right hand side is greater than 1.

Step 2: Recursive Formulation of the Loss We now show that the right hand side goes to 0 at an exponential rate. We define the loss function to be:

$$L_t(x_i) = \sum_{\tilde{y}\neq y_i} e^{\eta(F_{\tilde{y}}^T(x_i) - F_{y_i}^T(x_i))}, \tilde{L}_t = \frac{1}{m}\sum_{i=1}^{m} L_t(x_i).$$

We first express $\tilde{L}_{t+1}$ as a function of $\tilde{L}_t$. For any observation that is classified correctly at the $t+1$th stage, we multiply that observation's loss by a factor of $e^{-\eta}$. On the other hand, for any observation that is classified incorrectly as $\tilde{y}$, we *add* the following:

$$e^{\eta(F_{\tilde{y}}^t(x_i) - F_{y_i}^t(x_i))}(e^{\eta}-1).$$

So:

$$\tilde{L}_{t+1} = \frac{1}{m}\left(\sum_{i:h_{t+1}(x_i)=y_i} e^{-\eta}L_t(x_i) + \sum_{i:h_{t+1}(x_i)\neq y_i} L_t(x_i) + e^{\eta(F_{h_{t+1}(x_i)}^t(x_i) - F_{y_i}^t(x_i))}(e^{\eta}-1)\right).$$

Step 3: Weak Learnability By the above, h_{t+1} is chosen to solve:

$$\min_{h\in\mathcal{H}} \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}[h(x_i) = y_i]\left((e^{-\eta}-1)\sum_{\tilde{y}\neq y_i} e^{F_{\tilde{y}}^t - F_{y_i}^t}\right) + \mathbf{1}[h(x_i)\neq y_i](e^{\eta}-1)e^{\eta(F_{h(x_i)}^t - F_y^{t-1}(x_i))}.$$

In fact, using the previous step, we see that this can equivalently be expressed as $\tilde{L}_{t+1} - \tilde{L}_t$. On the other hand, someone who is random guessing, but is correct with extra probability γ, will be correct with probability $\frac{1-\gamma}{k} + \gamma$, and guess an incorrect label $\tilde{y}$ with probability $\frac{1-\gamma}{k}$. Furthermore, the hypothesis class ensures a weakly lower error (as measured by this cost) than the random guessing. Hence this expression is bounded above by:

$$\frac{1}{m}\sum_{i=1}^{m}\left((\frac{1-\gamma}{k}+\gamma)(e^{-\eta}-1)L_t(x_i)+\frac{1-\gamma}{k}\sum_{\tilde{y}\neq y_i}(e^{\eta}-1)e^{\eta(F^t_{\tilde{y}}(x_i)-F_y(x_i))}\right)$$

We now again see that the expression for $L_t(x_i)$ appears. Hence, rearranging, we obtain:

$$\left((\frac{1-\gamma}{k}+\gamma)(e^{-\eta}-1)+\frac{1-\gamma}{k}(e^{\eta}-1)\right)\tilde{L}_t.$$

Putting this together, we have the recursion:

$$\tilde{L}_{t+1}\leq\left(1+\left((\frac{1-\gamma}{k}+\gamma)(e^{-\eta}-1)+\frac{1-\gamma}{k}(e^{\eta}-1)\right)\right)\tilde{L}_t.$$

Step 4: Specifying η To complete the argument, we must specify an η which delivers the exponential convergence. However, first note that if $\eta=0$, the coefficient on $\tilde{L}_t$ in the previous inequality is 1, and the derivative with respect to η is $-\gamma$ at 0, so that this expression is less than 1, for some $\eta>0$. Setting $\eta=\log(1+\gamma)$, the above coefficient on $\tilde{L}_t$ reduces to:

$$1+\left((\frac{1-\gamma}{k}+\gamma)(\frac{1}{1+\gamma}-1)+\frac{1-\gamma}{k}\gamma\right)$$

Since this is bounded above by $e^{-\gamma^2/2}$, and since since $\tilde{L}_0=(k-1)$, we therefore have that:

$$\tilde{L}_t\leq(k-1)e^{-\gamma t^2/2},$$

as desired.

Appendix F. Proofs for Section 7.4

F.1. **Proof of Proposition 7.6.** The proof of the theorem proceeds in the following steps:

- Step 1: Show that the expected value conditional on price, in the image of the seller's possible strategies after applying the augmentation, is uniformly equicontinuous.
- Step 2: Show that the same label is applied to $\mathbf{E}[v_\theta \mid p+z_{i,\eta},\sigma,\phi_\eta]$ as would be applied to $\mathbf{E}[v_\theta \mid p,\sigma]$, with high probability.
- Step 3: Verify that the change in recommendation can be minimized uniformly

The condition that σ is either discrete or continuous is stronger than necessary; what is necessary is continuity of the conditional expectation as a function of price, which can be satisfied if the discrete portions and continuous portions are separated, for instance. However, the proposition highlights that we need not restrict the principal's strategy space at all in order for our algorithm to converge.

The Theorem implies that if the principal were to use an *arbitrary* strategy σ, the agent could instead focus on finding a rational response to $\tilde{\sigma}_\eta$. Doing so would still lead to PAC learnability of the approximately optimal response to σ. On the other hand, we can show that the optimal response to $\tilde{\sigma}_\eta$ is PAC learnable (unlike, potentially, the optimal response to σ), and doing the change leads to a negligible impact on the principal's surplus.

Before presenting the proof, we argue that uniform equicontinuity implies weak learnability. Suppose that $\mathbf{E}[v \mid \sigma,p]-p$ is uniformly equicontinuous (which holds if $\mathbf{E}[v \mid \sigma,p]$ is uniformly equicontinuous). By uniform equicontinuity, we have there exists some δ such that whenever $\left|p-p'\right|<\delta$, we have that

$$\left|\mathbf{E}[v \mid \sigma,p]-\mathbf{E}[v \mid \sigma,p']\right|<2\varepsilon,$$

for any σ. Suppose we have some price p such that $\mathbf{E}[v \mid \sigma,p]-p>\varepsilon$. Then if $\mathbf{E}[v \mid \sigma,p']-p'<-\varepsilon$, it follows that $\left|p-p'\right|>\delta$. It follows that there can only be at most $\frac{v_H-v_L}{\delta}$ prices such that $y(\sigma,p)=-y(\sigma,p')$, where p and p' are adjacent (ignoring all prices where $\left|\mathbf{E}[v \mid \sigma,p]-p\right|<\varepsilon$, as the classification decision is irrelevant there).

F.1.1. *Step One.* We first show that $\mathbf{E}[v_\theta \mid \tilde{\sigma}_\eta, p]$ is Lipschitz in p uniformly of $\tilde{\sigma}_\eta$, noting that we are restricting to prices where $\tilde{\sigma}_\eta(p) > \gamma$. Note that:

$$\tilde{\sigma}'_\eta(p \mid \theta) = \int \phi'_\eta(p - \tilde{p})\sigma(\tilde{p} \mid \theta)d\tilde{p} \leq \max \phi'_\eta := \overline{\phi'}.$$

Furthermore, we have:

$$\frac{d}{dp}\mathbf{P}_{\tilde{\sigma}_\eta}[\theta \mid p] = \frac{\tilde{\sigma}'_\eta(p \mid \theta)\mathbf{P}[\theta]}{\sum_{\tilde{\theta}} \tilde{\sigma}_\eta(p \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}]} - \frac{\tilde{\sigma}_\eta(p \mid \theta)(\sum_{\tilde{\theta}} \sigma'_\eta(p \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}])}{(\sum_{\tilde{\theta}} \tilde{\sigma}_\eta(p \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}])^2},$$

so:

$$\left|\frac{d}{dp}\mathbf{P}_{\tilde{\sigma}_\eta}[\theta \mid p]\right| \leq \overline{\phi'}\mathbf{P}[\theta] \cdot \left(\frac{1}{\sum_{\tilde{\theta}} \tilde{\sigma}_\eta(p \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}]}\right) + \overline{\phi'}\left(\frac{\tilde{\sigma}_\eta(p \mid \theta)\mathbf{P}[\theta]}{(\sum_{\tilde{\theta}} \tilde{\sigma}_\eta(p \mid \tilde{\theta}\mathbf{P}[\tilde{\theta}]))^2}\right) \leq \overline{\phi'}\left(\frac{\mathbf{P}[\theta]}{\gamma} + \frac{1}{\gamma}\right)$$

Hence we see that for all $p \neq p^*$, the conditional probability has a uniformly bounded derivative, and is hence Lipschitz continuous. Importantly, the bound only depends on η and γ (and $\mathbf{P}[\theta]$), and is therefore uniform over all strategies in the image of the augmentation. Hence we can ensure that Lipschitz continuity is mainted for all prices in the support of $\tilde{\sigma}_\eta$.

In fact, recall that the Lipschitz constant is equal to the L^∞ norm of the derivative. Hence Lipschitz continuity depends only on γ and $\overline{\phi'_\eta}$, meaning that the Lipschitz constant holds uniformly over the image of the distributions emerging under the algorithm. It follows that the image is uniformly equicontinuous.

F.1.2. *Step Two.* Note that since $\mathbf{E}[v_\theta \mid \sigma, p]$ is continuous on $S = \cup_\theta \text{ Supp } \sigma(\cdot \mid \theta)$, $\mathbf{E}[v_\theta \mid \sigma, p]$ is uniformly continuous on any compact $K \subset S$. Define:

$$K_\gamma = \{p : \sum_\theta \sigma(p \mid \theta)\mathbf{P}[\theta] \geq \gamma\}.$$

Using that mollifiers converge uniformly on compact sets, we have that $\tilde{\sigma}_\eta \to \sigma$ uniformly on K_γ. We therefore have that, for any $\tilde{\varepsilon}$, we can find some $\overline{\eta}$ such that if $\eta < \overline{\eta}$ and $p \in K_\gamma$, then $\left|\tilde{\sigma}_\eta(p \mid \theta) - \sigma(p \mid \theta)\right| < \tilde{\varepsilon}$ for all θ, and $\left|\sum_\theta \tilde{\sigma}_\eta(p \mid \theta)\mathbf{P}[\theta] - \sum_\theta \sigma(p \mid \theta)\mathbf{P}[\theta]\right| < \tilde{\varepsilon}$.

Furthermore, since σ is uniformly continuous on K_γ, we have:

$$\left|\sigma(p \mid \theta) - \tilde{\sigma}(p' \mid \theta)\right| = \left|\int \phi_\eta(p' - \tilde{p})(\sigma(p \mid \theta) - \sigma(\tilde{p} \mid \theta))d\tilde{p}\right| \leq \tilde{\varepsilon},$$

using the uniform continuity of σ on K_γ.

So for any $p \in K_\gamma$, and η sufficiently small, we have (letting $\overline{v} = \max_\theta v_\theta$):

$$\begin{aligned}
\left|\mathbf{E}[v_\theta \mid \sigma, p] - \mathbf{E}[v_\theta \mid \tilde{\sigma}_\eta, p']\right| &= \left|\frac{\sum_\theta v_\theta \sigma(p \mid \theta)\mathbf{P}[\theta] \sum_{\tilde{\theta}} \tilde{\sigma}_\eta(p' \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}] - \sum_\theta v_\theta \tilde{\sigma}_\eta(p' \mid \theta)\mathbf{P}[\theta] \sum_{\tilde{\theta}} \sigma(p \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}]}{\left(\sum_\theta \sigma(p \mid \theta)\mathbf{P}[\theta]\right)\left(\sum_\theta \tilde{\sigma}_\eta(p' \mid \theta)\mathbf{P}[\theta]\right)}\right| \\
&\leq \frac{1}{\sigma(p)\cdot(\gamma - \tilde{\varepsilon})}\left|\sum_\theta v_\theta(\sigma(p \mid \theta) - \tilde{\sigma}_\eta(p' \mid \theta))\mathbf{P}[\theta] \sum_{\tilde{\theta}} \sigma(p \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}]\right. \\
&\quad \left. + \sum_\theta v_\theta \sigma(p \mid \theta)\mathbf{P}[\theta] \sum_{\tilde{\theta}} (\tilde{\sigma}_\eta(p' \mid \tilde{\theta}) - \sigma(p \mid \tilde{\theta}))\mathbf{P}[\tilde{\theta}]\right| \\
&\leq \frac{1}{\sigma(p)\cdot(\gamma - \tilde{\varepsilon})}\left(\overbrace{\left|\sum_\theta v_\theta(\sigma(p \mid \theta) - \tilde{\sigma}_\eta(p' \mid \theta)) \sum_{\tilde{\theta}} \sigma(p \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}]\right|}^{\leq \overline{v}\tilde{\varepsilon}\sigma(p)}\right. \\
&\quad \left. + \overbrace{\left|\sum_\theta v_\theta \sigma(p \mid \theta)\mathbf{P}[\theta] \sum_{\tilde{\theta}} (\tilde{\sigma}_\eta(p' \mid \tilde{\theta}) - \sigma(p \mid \tilde{\theta}))\mathbf{P}[\tilde{\theta}]\right|}^{\leq \overline{v}\cdot\tilde{\varepsilon}\cdot\sigma(p)}\right) \\
&\leq \frac{2\overline{v}\tilde{\varepsilon}}{\gamma - \tilde{\varepsilon}}.
\end{aligned}$$

The second inequality follows from adding and subtracting $\sum_\theta v_\theta \sigma(p \mid \theta) \sum_{\tilde{\theta}} \sigma(p \mid \tilde{\theta})\mathbf{P}[\tilde{\theta}]$ to both sums inside the absolute value, and the second inequality is from the triangle inequality, and the overbraced expression follows from $v_\theta \leq \overline{v}$ and uniform convergence of $\tilde{\sigma}_\eta$ to σ.

So for any fixed γ, we can find some some η such that whenever $\eta < \overline{\eta}$, we can ensure that on K_γ, $\left|\mathbf{E}[v_\theta \mid \tilde{\sigma}_\eta, p] - \mathbf{E}[v_\theta \mid \sigma, p]\right| < \varepsilon^*$, by choosing $\tilde{\varepsilon}$ sufficiently small so that $\frac{2\tilde{\varepsilon}}{\gamma(\gamma - \tilde{\varepsilon})} < \varepsilon^*$. It follows that if the buyer's classifier converges to a rule that is ε-optimal under $\tilde{\sigma}_\eta$, it converges to a rule that is $\varepsilon + \varepsilon^*$ optimal under σ. The probability that this fails to occur is simply the probability that the price is outside of K_γ, which can be made arbitrarily small by taking $\gamma \to 0$, since we can approximate the support of σ arbitrarily well.

F.1.3. *Step Three.* Note that, for an arbitrary continuous distribution f, if $p \sim f$ we have (for any compact K):

$$\mathbf{P}_f[L_\gamma] = \int_K \mathbf{1}[p : f(p) \leq \gamma] f(p) dp \leq \int_K \mathbf{1}[p : f(p) \leq \gamma] \gamma dp \leq \mu(K) \cdot \gamma,$$

where μ is Lebesgue measure. It follows that the probability that $p \in L_\gamma$, is small if γ is small, and furthermore that this probability can be made small uniformly, using only γ.

As shown by the claim above, by taking η small, we can ensure that the difference in the conditional expected value is small with high probability. By taking γ small, we ensure that the probability of a different outcome due to smoothing goes to 0, implying the result.

References

Al-Najjar, N. I. (2009): "Decision Makers as Statisticians: Diversity, Ambiguity and Learning," *Econometrica*, 77(5), 1371–1401.

Al-Najjar, N. I., and M. M. Pai (2014): "Coarse decision making and overfitting," *J. Economic Theory*, 150, 467–486.

Blum, A., M. Hajiaghayi, K. Ligett, and A. Roth (2008): "Regret minimization and the price of total anarchy," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pp. 373–382.

Braverman, M., J. Mao, J. Schneider, and M. Weinberg (2018): "Selling to a No-Regret Buyer," in *ACM Conf. on ACM Conference on Economics and Computation (ACM EC)*, pp. 523–538.

Cherry, J., and Y. Salant (2019): "Statistical Inference in Games," Northwestern University.

Cho, I.-K., and K. Kasa (2015): "Learning and Model Validation," *Review of Economic Studies*, 82, 45–82.

Deng, Y., J. Schneider, and B. Sivan (2019): "Strategizing against No-regret Learners," Discussion paper.

Dietterich, T. G. (2000): "Ensemble Methods in Machine Learning," in *Multiple Classifier Systems*, pp. 1–15, Berlin, Heidelberg. Springer Berlin Heidelberg.

Eliaz, K., and R. Spiegler (2018): "A Model of Competing Narratives," Brown University and Tel Aviv University.

Esponda, I., and D. Pouzo (2014): "An Equilibrium Framework for Players with Misspecified Models," University of Washington and University of California, Berkeley.

Fudenberg, D., and K. He (2018): "Learning and Type Compatibility in Signaling Games," *Econometrica*, 86(4), 1215–1255.

Fudenberg, D., and D. M. Kreps (1995): "Learning in Extensive Form Games I: Self-confirming Equilibria," *Journal of Economic Theory*, 8(1), 20–55.

Fudenberg, D., and D. K. Levine (1993): "Steady State Learning and Nash Equilibrium," *Econometrica*, 61(3), 547–573.

——— (2006): "Superstition and Rational Learning," *American Economic Reivew*, 96, 630–651.

Hornik, K., M. Stinchcombe, and H. White (1989): "Multilayer feedforward networks are universal approximators," *Neural Networks*, 2(5), 359 – 366.

Liang, A. (2018): "Games of Incomplete Information Played by Statisticians," Discussion paper, University of Pennsylvania.

Marcet, A., and T. J. Sargent (1989): "Convergence of Least Squares Learning Mechanisms in Self Referential Linear Stochastic Models," *Journal of Economic Theory*, 48, 337–368.

Maskin, E., and J. Tirole (1992): " The Principal-Agent Relationship with an Informed Principal, II: Common Values," *Econometrica*, 60(1), 1–42.

Mukherjee, I., and R. E. Schapire (2013): "A Theory of Multiclass Boosting," *Journal of Machine Learning Research*, 14, 437–497.

Nekipelov, D., V. Syrgkanis, and E. Tardos (2015): "Econometrics for Learning Agents," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 1–18.

Olea, J. L. M., P. Ortoleva, M. M. Pai, and A. Prat (2019): "Competing Models," Columbia University, Princeton University and Rice University.

Rubinstein, A. (1986): "Finite Automata Play Repeated Prisoners Dilemma," *Journal of Economic Theory*, 39(1), 83–96.

——— (1993): "On Price Recognition and Computational Complexity in a Monopolistic Model," *Journal of Political Economy*, 101(3), 473–484.

Schapire, R. E., and Y. Freund (2012): *Boosting: Foundations and Algorithms*. MIT Press.

Spence, A. M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87(3), 355–374.

Spiegler, R. (2016): " Bayesian Networks and Boundedly Rational Expectations *," *The Quarterly Journal of Economics*, 131(3), 1243–1290.

Zhao, C., S. Ke, Z. Wang, and S.-L. Hsieh (2020): "Behavioral Neural Networks," Discussion paper.

Department of Economics, Emory University, Atlanta, GA 30322 USA
E-mail address: icho30@emory.edu
URL: https://sites.google.com/site/inkoocho

Department of Economics, University of Southern California, Los Angeles, CA 90089 USA
E-mail address: libgober@usc.edu
URL: http://www.jonlib.com/