Statistical Approach to Robust Mechanism Design

Modibo K. Camara[*]

September 21, 2021

Abstract

In many settings, understanding an agent's beliefs can be the difference between a successful policy and one that fails dramatically. However, the agent's beliefs may not be identified by the available data. In this paper, I propose a modeling assumption that bypasses this issue, and show that it can highlight important policy tradeoffs not captured by existing models. The core idea is simple: if the available data convincingly demonstrates some fact about the world, the agent should believe that fact. Otherwise, her beliefs are left unspecified.

I develop this approach in the context of incomplete-information games where a policy-maker commits to a policy, an agent responds, and both access a public dataset. Policies that are too complex may be suboptimal because they lead to unpredictable behavior. To balance the benefits of policy complexity with its costs, I develop a method called *strategic regularization* and motivate it through both theoretical guarantees and illustrative examples.

[*] Department of Economics, Northwestern University. Email: mcamara@u.northwestern.edu.

Contents

1 Introduction

It is a truism in economics that beliefs are important determinants of behavior. In any number of settings, ranging from vaccine distribution to compensation policy, understanding agent's beliefs can make the difference between a successful policy and one that fails dramatically. Unfortunately, in many instances, the rich behavioral or survey data needed to identify agents' beliefs may not be available. This can limit our ability to provide concrete policy recommendations.

While existing modeling techniques can circumvent this lack of identification, they are widely recognized as imperfect. Let us consider three prominent techniques. First, the common prior assumption posits that the agent shares the policymaker's beliefs, whatever they may be. This has been criticized within economic theory as unrealistic, as part of the Wilson doctrine (e.g. Bergemann and Morris 2005). Second, the rational expectations assumption posits that the agent knows the true data-generating process. This has been criticized, in particular, because it assumes agents can somehow bypass inferential issues that plague empirical research (Manski 1993, Manski 2004a). Third, robust approaches do not make any assumptions on the agent's beliefs; instead, they optimize against worst-case beliefs using criteria like minimax regret. Robust approaches are very credible, but their recommendations are often too conservative to be useful in practice.

In this paper, I propose a new technique and show that it can highlight important policy tradeoffs that are not captured by existing alternatives. The core idea is straightforward: agent's beliefs are disciplined by the available data, but otherwise left unspecified. If the data convincingly demonstrates some fact about the world, the agent should believe that fact. On the other hand, if there is insufficient data to reach a particular conclusion, the agent's beliefs are undetermined. Critically, policy choices affect both the conclusions that the agent needs to evaluate, and how much data is needed to determine their validity. Policies that are too complex for the agent, in a sense to be made precise, may be suboptimal because they lead to unpredictable behavior.

To balance the benefits of policy complexity with its costs, I develop a method called *strategic regularization* and motivate it through both theoretical guarantees and illustrative examples. This method produces data-driven policies for incomplete information games where a policymaker commits to a policy, an agent responds, and both have access to a public dataset. Roughly, strategic regularization optimizes against worst-case agent behavior, subject to the constraint that the agent's beliefs are disciplined by the available data. My theoretical guarantees show that we can estimate the strategically-regularized policy, and characterize the rate of convergence. My illustrative examples show that strategic regularization can lead to new insights in models of vaccine distribution, prescription drug approval, performance pay, and product bundling.

Model. I consider a rich class of incomplete information games where a policymaker commits to a policy and a single agent responds. Payoffs are determined in part by a hidden state of nature. The state is drawn from some unknown distribution. This model captures a number of design problems, like monopoly regulation, Bayesian persuasion, and contract design. For example, take contract design. There, the policy maps observed performance to wages, the agent's response is her hidden action, and the state of nature maps the hidden action to observed performance.

Next, I assume that both the policymaker and the agent have access to a public dataset. This dataset consists of n i.i.d. draws from the state distribution. The strategies of both participants are statistical decision rules, mapping the realized sample to actions. In other words, both participants face statistical decision problems (see e.g. Wald 1950, Manski 2004b). However, we will see that neither the policymaker's nor the agent's problem is entirely standard.

Behavioral Assumptions. My goal is to recommend a strategy to the policymaker that performs well, according to his objective, under reasonable assumptions on the agent's behavior. To formalize these behavioral assumptions, I adapt ideas from statistical learning theory.

My approach involves bounding the agent's regret. Regret measures how suboptimal the agent's strategy is, in expectation, according to the true distribution. An agent that knows the true distribution can guarantee zero regret, so allowing the agent to obtain positive regret is a simple way to allow for misspecified beliefs. Of course, the agent's beliefs cannot be entirely unrestricted. They need to be disciplined by the available data. To capture this, I bound the agent's regret as a function of the sample size, realized policy, and state distribution.

A natural requirement is that the agent can always satisfy this regret bound, even if she lacks prior knowledge about the distribution. Essentially, this is a feasibility requirement. To formalize this, I turn to a off-the-shelf heuristic called *empirical utility maximization*. This heuristic is widely used in statistics and related fields, and does not require any prior knowledge or tuning parameters other than the agent's utility function. For these reasons, I call a regret bound feasible if the agent can satisfy that bound by using empirical utility maximization.[1]

To formulate a feasible regret bound, I adapt a central concept from statistical learning theory, called *Rademacher complexity*. This measures the complexity of the agent's learning problem. Naturally, an agent faced with a more complex learning problem should be more likely to accumulate regret, and this is reflected in my regret bound. Moreover, in this model, the agent's learning problem varies based on the policymaker's choice of policy. As such, Rademacher complexity depends on the chosen policy, and captures a form of policy complexity from the agent's perspective.

To complete the specification of the regret bound, I introduce a concept called *sample privacy* that is closely related to differential privacy (Dwork et al. 2006). This is a property of the policymaker's strategy, and controls how aggressively the policymaker makes use of the realized sample. The reason why sample privacy is needed is a bit subtle. The agent is trying to learn an optimal response to the chosen policy, using the realized sample. But the chosen policy also depends on the realized sample. Essentially, the agent is trying to learn an optimal response to a moving target. If the policymaker commits to using the realized sample less aggressively, as measured by sample privacy, the target moves less and the agent accumulates less regret.

Strategic Regularization. This model introduces new trade-offs. Policies that are more complex or more sensitive to the data mean looser regret bounds, and hence, a wider range of behaviors that

[1] To be clear, I do not assume that the agent uses empirical utility maximization. However, I do find it reasonable to assume that the agent's strategy does not consistently underperform empirical utility maximization. Otherwise, she would have been better off replacing her strategy.

the agent can engage in. In the worst case, less predictable behavior means worse outcomes for the policymaker. To make the agent's behavior more predictable, the policymaker can choose policies that are less complex and less sensitive to the data. However, these changes can be costly.

Strategic regularization balances these trade-offs. Like other uses of regularization in statistics, it biases the policymaker towards policies that are less complex and strategies that are more stable. It evaluates policies according to the policymaker's objective, under the worst-case agent response that is consistent with my regret bound. If there is enough data for the agent to determine the optimal response, the regret bound will be close to zero and the agent's response will be nearly optimal with high probability. Otherwise, the regret bound may be large and strategic regularization will guard against suboptimal responses driven by misspecified beliefs.

My first theorem shows that it is possible to estimate the optimal strategically-regularized policy. Like empirical utility maximization, the proposed estimator evaluates the policymaker's expected payoff with respect to the empirical distribution. It then estimates the set of behaviors that satisfy the agent's regret bound, and optimizes against the worst case behavior in this estimated set. The key challenge in constructing this estimator has to do with sample privacy. Sample privacy is a property of the estimator itself, but it also enters into the regret bound, which I need to define the estimator in the first place. I resolve this circularity and guarantee sample privacy by using the exponential mechanism (McSherry and Talwar 2007), with appropriately tuned parameters.

My second theorem shows that the proposed estimator's rate of convergence is approximately optimal. I begin by showing that the policymaker's expected payoff converges to his optimal common knowledge payoff, where both participants know the true distribution. Next, I evaluate the rate of convergence. The first-best rate is obtained by the optimal strategically regularized policy, where the policymaker knows the distribution but the agent is still learning. If the first-best rate is $n^{-\gamma}$, I show that the proposed estimator's rate of convergence is at least $n^{-\frac{\gamma}{1+2\gamma}}$. In typical applications where $\gamma = 1/2$, my estimator's rate of convergence is $n^{-1/4}$.

Illustrative Examples. I argue strategic regularization can lead to new insights in four examples. These examples are not intended to be as general or realistic as possible. Instead, they are meant to convey a core insight that motivates the use of strategic regularization in similar applications.

First, I consider a model of vaccine distribution. Here, strategic regularization suggests waiting for statistically-significant clinical trial results before distributing vaccines. While this is common practice, it has been criticized (e.g. Wasserstein and Lazar 2016) and conflicts with the recommendations of the treatment choice literature (e.g. Manski 2019). My model justifies this practice. People may not take up the vaccine if it has not been proven effective beyond a reasonable doubt. If there are fixed costs to vaccine distribution, this outcome is worse than not distributing the vaccine in the first place. So, while there is no statistical reason to insist on statistical significance per se, there is a strategic reason. Namely, the population needs to be convinced of vaccine quality.

Second, I consider a model of prescription drug approval by a regulator. Here, strategic regularization restricts doctors' ability to prescribe drugs that haven't been proven effective in clinical trials. The standard for approval increases as more drugs are approved, as in stepwise methods for multiple hypothesis testing (e.g. Holm 1979, Romano and Wolf 2005). In contrast, in models with a common prior or rational expectations, the optimal policy is to approve all drugs. Essentially, this delegates the decision to doctors, who are better informed than the regulator. In my model, however, doctors may prescribe ineffective drugs if the clinical trial returns a "false positive" where an ineffective drug appears to be effective by random chance. Limiting the number of drugs approved can reduce the risk of false positives, and provide better welfare guarantees.

Third, I consider a model of performance pay. An employer incentivizes an employee to exert costly effort by paying wages contingent on observed performance. Here, strategic regularization caps and flattens the wage schedule. Wages are zero until they reach some threshold level of performance. At that point, they jump and remain flat. When the sample size is small, the maximum wage is relatively small and easy to obtain. In contrast, the common prior solution compensates the employee only when the maximum possible performance is obtained, and pays a very large wage. In my model, this does not work well. If historical data is limited, it may not be obvious to the employee that it is worth investing effort for a small chance of receiving the bonus. Flatter contracts may be less potent, but they provide clearer incentives.

Finally, I consider a model of product bundling. A firm has several products for sale and wants to sell them in a way that maximizes expected profit. Here, strategic regularization favors selling large bundles of products, or even bundling all products together into a grand bundle. This contrasts with prior work that suggests selling all items separately is robustly optimal, when there is not much information about consumer demand (Carroll 2017). In my model, the reason for bundling is that consumers learn about their value for the product through reviews. If there are many products, but few reviews per product, consumers can be confident in the value of the grand bundle while being uncertain about the value of any given product. In that case, all else equal, it is easier to convince consumers to buy the bundle.

Related Literature This work contributes to three research efforts. For robust mechanism design, it is a principled way to interpolate between two extremes: the common prior and prior-freeness. This goal is similar to Artemov et al. (2013) and Ollár and Penta (2017). For learning in games, I provide a convenient behavioral assumption that does not rely on agents using a particular estimator. The most similar model is that of Liang (2020). For data-driven mechanism design, I extend existing work to settings where the agent is learning, not just the policymaker. This is also true of Camara et al. (2020), Cummings et al. (2020), and Immorlica et al. (2020). I leave a more detailed discussion of the prior literature to section 6.

Organization. I begin in section 2 by introducing the model. Section 3 formalizes the behavioral assumptions, with Rademacher complexity defined in subsection 3.1 and sample privacy in subsection 3.2. Section 4 contains the main results. I define strategic regularization, develop an estimator in subsection 4.1, and characterize its convergence in subsection 4.2. In section 5, I present the four illustrative examples. Section 6 discusses related literature and section 7 concludes. Omitted proofs can be found in appendix A.

2 Model

This paper studies Stackelberg games of incomplete information. There are two players: a policymaker and an agent. The policymaker moves first and commits to a policy $p \in \mathcal{P}$, or possibly a mixed policy $\pi^p \in \Delta(\mathcal{P})$. After observing the policy p, the agent chooses a response $r \in \mathcal{R}$, or possibly a mixed response $\pi^r \in \Delta(\mathcal{R})$. Finally, the state of nature $s \in \mathcal{S}$ is drawn from a distribution $\pi^s \in \Delta(\mathcal{S})$. The policymaker's utility is given by $u^P(p, r, s)$ and the agent's utility by $u^A(p, r, s)$.

This setup is quite general. It can be used to model a variety of single-agent problems in mechanism design, contract design, information design, and other areas of interest to economists. Features like transfers, messages between the policymaker and agent, and asymmetric information about the state can be included (or excluded) by defining the policy space $\mathcal{P}$ and the response space $\mathcal{R}$ appropriately.[2] The only really essential feature is that the policymaker can commit to a policy before the agent responds. I also maintain the following assumptions throughout the paper.

Assumption 1. *The following assumptions hold:*

1. *The policy space $\mathcal{P}$ is finite, with $n_{\mathcal{P}} < \infty$ elements.*

2. *For each participant $i \in \{P, A\}$, the utility function u^i is bounded. It follows that the maximum variation in u^i given a policy p is finite, i.e.*

$$\Delta^i(p) = \sup_{r,s} u^i(p, r, s) - \inf_{r,s} u^i(p, r, s) < \infty$$

The reader is welcome to think of the policy space as a discrete approximation to an infinite policy space. In fact, this is what I do in two of the illustrative examples in section 5.

As discussed in the introduction, it is common to assume that both the policymaker and the agent know the distribution π^s. In this case, the policymaker's problem is entirely standard. The agent is free to choose any response r that maximizes her expected utility, i.e.

$$r \in \arg\max_{r'} \mathrm{E}_{\pi^s}\left[u^A(p, r', s)\right]$$

In turn, the policymaker chooses a policy p that maximizes his expected utility after taking into account how the agent will respond. Of course, the policymaker may not know how the agent chooses between responses she is indifferent to. However, the policymaker can optimize against the worst-case response, and guarantee an arbitrarily close approximation to the following payoff:

$$\begin{aligned} \mathrm{CK}(\pi^s) = \max_p \min_r \; & \mathrm{E}_{\pi^s}\left[u^P(p, r, s)\right] \qquad (1) \\ \text{s.t.} \quad & r \in \arg\max_{r'} \mathrm{E}_{\pi^s}\left[u^A(p, r', s)\right] \end{aligned}$$

[2] For example, if the policymaker has access to transfers, then the policy p must specify the transfers made (if any). Alternatively, if the agent has private information about some aspect of the state s, then her response r should map that information to some action. Finally, if the agent reports her private information to the policymaker, then p should map that report to some action.

Here, CK stands for common knowledge.[3]

The common knowledge assumption has two obvious drawbacks. The first is that the policymaker may not know the true distribution π^s. It seems self-evident that if we recommend a policy to the policymaker, it should not rely on information that he does not have access to. The second drawback is that the agent may not know the distribution. In that case, our predictions about how she responds to a given policy may be incorrect. This can lead to suboptimal policy choices.

Next, I introduce a framework that avoids these two drawbacks. It does so by replacing common knowledge with data, and introducing new, weaker assumptions on the agent's behavior.Specifically, I assume that that both the policymaker and the agent can learn about the distribution π^s from a publicly-available dataset. A dataset consists of n i.i.d. observations of the state, i.e.

$$S_1, \dots, S_n \sim \pi^s$$

Each participant's strategy is now a statistical decision rule. Specifically, the policymaker's strategy maps the dataset to a distribution over policies, i.e.

$$\sigma_n^D : \mathcal{S}^n \to \Delta(\mathcal{P})$$

The agent's strategy maps the dataset and the policymaker's policy to a distribution over responses, i.e.

$$\sigma_n^A : \mathcal{S}^n \times \mathcal{P} \to \Delta(\mathcal{R})$$

Let $P_n \sim \sigma_n^D(S_1, \dots, S_n)$ denote the realized policy, and let $R_n \sim \sigma_n^A(S_1, \dots, S_n, P_n)$ denote the realized response. These are random variables, both because the participants may use mixed strategies and because they depend on the underlying random sample.

It is worth emphasizing that this dataset is ideal in several ways. First, the states of nature S_i are directly observed. This is a good starting point because the data clearly identifies the distribution π^s. In practice, however, the state of nature may not be observed directly. In these cases, the distribution π^s may only be partially identified. Second, the observations S_i are drawn independently from the true distribution π^s. This assumption is standard, but may fail in dynamic environments where historical data does not fully reflect the present. Third, the dataset is publicly available to both participants. More precisely, any data that the policymaker uses must also be available to the agent (it is not a problem if the agent has access to additional data). In some cases, the policymaker may be able to guarantee that this assumption holds by sharing his data with the agent.

[3] The reader may note that it is common to assume that the agent breaks ties (between responses that she is indifferent to) in favor of the policymaker. Here, I do not make any assumption about how the agent breaks ties. Instead, I have the policymaker optimize against the worst case. It is important to keep in mind that this often does not make a difference. For example, in mechanism design with transfers, it is often possible to tweak the transfers slightly and break the agent's indifference in the policymaker's favor. If so, the supremum over all policies p under worst-case tie-breaking will be the same as the maximum over all policies p under best-case tie-breaking. For that reason, best-case tie-breaking is often viewed as an innocuous assumption that is convenient because it ensures the existence of optimal policies. However, it is easy to construct games where best-case tie-breaking is far from innocuous. And it will become clear in proposition 3 that worst-case tie-breaking is more convenient for the purposes of this paper.

These caveats aside, my goal is to construct a strategy σ_n^D that will perform well according to the policymaker's utility function, given reasonable assumptions on the agent's strategy σ_n^A. Previously, I required the agent to maximize expected utility with respect to the true distribution π^s. In the next section, I will assume that the agent's strategy is not too suboptimal with respect to the true distribution, and that it becomes closer to optimal as more data becomes available (n increases). Beyond this, I do not require the policymaker to have any knowledge of the agent's strategy.

3 Behavioral Assumptions

In order to make policy recommendations, I need to impose rationality assumptions for the agent. My approach involves bounding the agent's *regret*. Allowing the agent to obtain positive regret is a relatively tractable – but still principled – way to allow for misspecified beliefs. Bounding the agent's regret limits how misspecified those beliefs can be, as a function of the sample size, realized policy, and state distribution.

Essentially, the agent's regret captures how suboptimal her strategy is. It is the difference between her optimal expected utility and the expected utility she achieves by following her strategy, i.e.

$$\mathrm{Regret}_n^A\left(\sigma_n^A, \sigma_n^P, \pi^s\right) = \max_r \mathrm{E}_{\pi^s}\left[u^A(P_n, r, s)\right] - \mathrm{E}_{\pi^s}\left[u^A(P_n, R_n, s)\right]$$

To be clear, expectations are taken with respect to the state s, the realized sample $S_1, \dots, S_n$, and the participants' mixed strategies. And, as stated above, the random variables R_n and P_n are generated by the strategies σ_n^A and σ_n^P respectively.

It turns out that a refined notion of regret will be more useful in this paper. The reason is that the policymaker not only cares about whether the agent makes mistakes, but also about how those mistakes are correlated with the policies that he chooses. In other words, the policymaker cares about the agent's regret conditional on the policy $P_n = p$. Formally, the agent's *conditional regret* is

$$\mathrm{Regret}_n^A\left(\sigma_n^A, \sigma_n^P, \pi^s \mid P_n = p\right) = \max_r \mathrm{E}_{\pi^s}[u(p, r, s)] - \mathrm{E}_{\pi^s}\left[u(p, R_n, s) \mid P_n = p\right]$$

Note that the first expectation is not conditioned on $P_n = p$ because P_n is independent of $u(p, r, s)$.

My assumption is that the agent's conditional regret is bounded. The bound is left abstract for the moment, but will be specified shortly.

Assumption 2. *There exists a regret bound $B\left(\sigma_n^P, \pi^s, p\right)$ such that the agent's strategy σ_n^A satisfies*

$$\mathrm{Regret}_n^A\left(\sigma_n^A, \sigma_n^P, \pi^s \mid P_n = p\right) \leq B\left(\sigma_n^P, \pi^s, p\right)$$

for any policymaker's strategy σ_n^P and policy p.

Allowing the agent to obtain positive regret (or conditional regret) is a conceptually-simple way of relaxing the prior knowledge assumption. It does not require us to define a set of prior beliefs that the agent could possess, which usually involves commiting to a metric on the space of beliefs.

It neither presupposes that the agent is Bayesian, nor does it rule that out. It does not exclude the possibility that the agent has additional information about the distribution π^s that goes beyond the public dataset. It does not even require the agent to use the public data, if she can achieve similar or better performance by other means (e.g. personal experience or good judgement).

The main drawback of this approach is that the regret bound $B\left(\sigma_n^P, \pi^s, p\right)$ depends on several complicated parameters. While much of this paper is devoted to addressing this drawback, it also turns out to be a blessing in disguise. The reason is that it highlights which features of a policy p tend to increase the agent's regret, and therefore make her behavior less predictable. As the examples in section 5 will illustrate, this makes it easy to identify features of policies that make them more – or less – robust to imperfect knowledge.

Next, I refine assumption 2 by specifying the function $B(\cdot)$. A natural requirement is that the agent can always satisfy this regret bound, even if she lacks prior knowledge about the distribution. Essentially, this is a feasibility requirement. To formalize this, I turn to a heuristic called *empirical utility maximization*, and adapt a concept called *Rademacher complexity*.

3.1 Rademacher Complexity

To formulate a feasible regret bound, I need a way to measure the complexity of the agent's learning problem. After all, not all learning problems are created equal.[4] Moreover, an agent faced with a more complex learning problem may be more likely to accumulate regret. Fortunately, I can adapt a central concept from statistical learning theory – Rademacher complexity – to provide a suitable measure. Since the Rademacher complexity will depend in part on the chosen policy, it will also capture a form of policy complexity from the agent's perspective.

I will define Rademacher complexity momentarily. But first, I need to introduce a well-known strategy called empirical utility maximization.

Definition 1. Empirical utility maximization $\hat{\sigma}_n^A$ *is a strategy for the agent. It chooses the response r that maximizes the agent's expected utility with respect to the empirical distribution* $\hat{\pi}^s$. *Formally,*

$$\hat{\sigma}_n\left(S_1, \dots, S_n\right) = \hat{R}_n = \arg\max_r \frac{1}{n}\sum_{i=1}^{n} u^A\left(P_n, r, S_i\right)$$

I consider a regret bound feasible if the agent can satisfy that bound by using empirical utility maximization. After all, this is a popular heuristic that is widely used in statistics and related fields (econometrics, machine learning, etc.). It would be too optimistic to assume that the agent in our model can outperform the state of the art. At the same time, empirical utility maximization is a simple and relatively tractable heuristic, which requires no prior knowledge about the distribution π^s.

[4] Intuitively, if the agent's utility function $u^A(p, r, s)$ involves intricate interactions between the response r and the state s, she may be more likely to choose suboptimal responses when the empirical distribution $\hat{\pi}^s$ is only slightly misspecified from the true distribution π^s. This would be a "complex" learning problem. This kind of complexity is often associated with response spaces $\mathcal{R}$ that are large or high-dimensional.In the other extreme, if the response space is a singleton, or the agent's utility $u^A(p, r, s)$ does not depend on the state at all, then she would always choose the optimal response. This does not require any learning.

The agent could always use this heuristic in lieu of her own strategy, and so should not consistently underperform empirical utility maximization.[5]

Rademacher complexity can be used to bound the agent's regret under empirical utility maximization. It measures the potential for this strategy to overfit to sampling noise. First, it trivializes the agent's learning problem by randomizing the sign of the agent's utility function. That is, it replaces the utility function $u^A(p, r, s)$ with $\sigma \cdot u^A(p, r, s)$, where $\sigma \sim \textsc{Uniform}\{-1, 1\}$ is known as a Rademacher random variable. This trivializes the learning problem since expected utility is zero for all responses r, i.e.

$$\mathrm{E}_{\pi^s}\left[\sigma \cdot u^A(p, r, s)\right] = 0$$

Second, Rademacher complexity asks how much the empirical utility maximizer $\hat{R}_n$ will overfit to this modified problem. Formally, the empirical utility is

$$\frac{1}{n}\sum_{i=1}^{n} \sigma_i \cdot u^A\left(p, r, S_i\right)$$

where σ_i are i.i.d. Rademacher random variables. For any given response r, the empirical utility is zero in expectation. On the other hand, the empirical utility of $\hat{R}_n$ will generally have a positive expected value in finite samples. This expected value is the Rademacher complexity, and reflects how severely the agent can be misled by the sampling noise in $\sigma_1, \dots, \sigma_n$.

Definition 2 (Bartlett and Mendelson 2003)**.** *The Rademacher complexity induced by policy p is*

$$\mathcal{RC}_n^A(p, \pi^s) = \mathrm{E}_{\pi^s}\left[\max_r \frac{1}{n}\sum_{i=1}^{n} \sigma_i \cdot u^A\left(p, r, S_i\right)\right]$$

where $\sigma_1, \dots, \sigma_n \sim \textsc{Uniform}\{-1, 1\}$ are i.i.d. Rademacher random variables.

Note that the Rademacher complexity will typically (but not always) converge to zero as $n \to \infty$. In other words, the agent's learning problem typically becomes easier as she obtains more data.

When the policymaker uses a constant strategy σ_n^P, the Rademacher complexity bounds the agent's regret. I will extend this result to non-constant strategies in proposition 2.

Proposition 1 (Bartlett and Mendelson 2003)**.** *Suppose the policymaker uses a constant strategy, where $\sigma_n^P(\cdot) = p$ for all arguments, and the agent follows the empirical utility maximizer $\hat{\sigma}_n^A$. Then the Rademacher complexity induced by the policy p bounds the agent's regret, where*

$$\mathrm{Regret}_n^A\left(\hat{\sigma}_n^A, p, \pi^s\right) \leq 4\mathcal{RC}_n^A(p, \pi^s) \tag{2}$$

This result does not hold for non-constant strategies σ_n^P. The regret bound for these strategies will inevitably be looser than the upper bound in (21), as the following example illustrates.

[5] In principle, I could rely on strategies other than empirical utility maximization to calibrate the regret bound $B(\cdot)$. For example, I could attempt to specify an uninformative prior distribution over distributions F and consider expected utility maximization with respect to that prior. Relative to this alternative, the advantage of empirical utility maximization is that it is easy to define and its properties are well-understood.

Example 1. Consider a model of prescription drug approval, which I expand in section 5.2. A population of patients is afflicted by a disease. There are m treatments available, and a placebo. Let ω_j be the average treatment effect of treatment j. The placebo's treatment effect is zero, and this is common knowledge. The patients are nonstrategic and accept whatever treatment is offered.

There is a regulator who approves treatments and a doctor who prescribes them. Formally, the regulator specifies a set $\mathcal{A} \subseteq \{1, \dots, m\}$ of approved treatments. Then the doctor either prescribes a treatment $j \in \mathcal{A}$ to a given patient, or prescribes the placebo. Both participants have access to clinical trial data with sample size n, where $\hat{\omega}_j$ is the sample ATE of treatment j. The doctor wants to maximize the patient's expected outcome, i.e. ω_j for the chosen treatment j.

The doctor's regret from expected utility maximization can vary significantly based on how the regulator uses the sample. Suppose the regulator approves a single treatment j independently of the sample. The doctor's regret will be on the order of $O\left(n^{-1/2}\right)$. Alternatively, suppose that regulator only approves the empirical utility maximizer, i.e. $\arg\max_j \hat{\omega}_j$. In that case, the doctor's choice is the same as if all treatments had been approved, and her regret will be on the order of $O\left(n^{-1/2} \log m\right)$. This dependence on m reflects a multiple testing problem that arises even though the agent only has two choices: the approved treatment and the placebo. It arises because the choice presented to her is correlated with the sample. □

To overcome the issues presented in this example, I introduce a property called *sample privacy*.

3.2 Sample Privacy

Sample privacy is a form of stability, in that it limits the dependence of the policy P_n on sampling noise. The definition is closely related to the notion of differential privacy, developed in Dwork et al. (2006) and subsequent work. When applied to a strategy σ_n^P, differential privacy ensures that the chosen policy P_n does not change much when any one sample point S_i is removed from the sample, without replacement. In contrast, sample privacy ensures that the policy P_n does not change much when the entire sample $S_1, \dots, S_n$ is dropped and replaced with a new sample $S_1', \dots, S_n' \sim \pi^s$. In that sense, the policy P_n can reflect knowledge about the distribution π^s, but not reveal much else about the noisy sample realizations $S_1, \dots, S_n$ that produced said knowledge.

Definition 3. *The policymaker's strategy σ_n^P satisfies* sample privacy with parameters (ϵ, δ) *if there exists an event $U \subseteq \mathcal{S}^n$ with probability exceeding $1 - \delta$ such that*

$$\Pr_{\pi^s}\left[P_n = p \mid S_1, \dots, S_n\right] \le e^{\epsilon} \cdot \Pr_{\pi^s}\left[P_n = p, U\right]$$

for any sample realizations $\left(S_1, \dots, S_n\right) \in U$.

Every strategy σ_n^P can be associated with a minimal privacy parameters $\epsilon, \delta \in \mathbb{R}_+ \cup \{\infty\}$, although it may be infinite. Constant strategies reflect an extreme case where $\epsilon, \delta = 0$. I discuss how the policymaker can achieve intermediate levels of sample privacy in section 3.1.

I can use the privacy parameters to generalize the regret bound in proposition 1. That bound only applied when the policymaker was using a constant strategy, or equivalently, a strategy with privacy parameters $\epsilon, \delta = 0$. The next proposition shows that this bound degrades smoothly as ϵ, δ increase.

Proposition 2. *Suppose the policymaker's strategy σ_n^P has privacy parameters (ϵ, δ), and the agent follows the empirical utility maximizer $\hat{\sigma}_n^A$. Then the Rademacher complexity induced by the policy p bounds the agent's regret conditional on $P_n = p$, where*

$$\text{Regret}_n^A\left(\hat{\sigma}_n^A, \sigma_n^P, \pi^s \mid P_n = p\right) \leq 4e^{\epsilon} \cdot \mathcal{RC}_n^A(p, \pi^s) + \delta \cdot \Delta^A(p) \tag{3}$$

Finally, I can revisit assumption 2 and specify the bound $B(\cdot)$ on the agent's regret. To recap, proposition 2 shows that an agent can guarantee the regret bound (3) by following the simple heuristic called empirical utility maximization. It stands to reason that a self-interested, sophisticated agent should not consistently obtain a higher regret than what she could achieve with this simple heuristic. As such, I assume that the agent's regret does not exceed the upper bound in (3).

Assumption 2 (continued)**.** *Let $\delta \geq 0$ be the privacy parameter of the policymaker's strategy σ_n^P. The agent's regret bound is given by*

$$B\left(\sigma_n^P, \pi^s, p\right) = 4e^{\delta} \cdot \mathcal{RC}_n^A(p, \pi^s) + \delta \cdot \Delta^A(p)$$

This completes the description of my model. The novel trade-off is that different strategies from the policymaker can mean more or less predictable behavior from the agent. –The policymaker can make the agent's behavior more predictable by choosing policies that are less complex, as measured by the Rademacher complexity, and by not overfitting to the data, as measured by sample privacy. As the regret bound $B(\cdot)$ increases, the agent's behavior becomes less predictable. In particular, the agent can make unpredictable mistakes that are costly for the policymaker. On the other hand, the policymaker has two ways of tightening the agent's regret bound. First, he can choose policies p that are less complex, as measured by Rademacher complexity. Second, he can use strategies that are less sensitive to sampling noise, as measured by sample privacy. However, both of these measures can be costly in their own right. –In general, these measures are *with* loss of optimality.

Next, I propose a robust approach that balances the advantages of complexity with the disadvantages of unpredictability. I call this approach *strategic regularization*.

4 Strategic Regularization

In this section, I formalize strategic regularization as a robust optimization problem. Then I present the main results of the paper. Theorem 1 shows that it is possible to estimate the optimal strategically-regularized policy. Theorem 2 characterizes the proposed estimator's rate of convergence. In the process, I introduce one more regularity assumption.

Strategic regularization biases the policymaker towards policies that are less complex and strategies that are more stable. In that sense, it is similar to other forms of regularization used in statistics. But in another sense, it is quite different. Regularization is usually intended to simplify one's own learning problem. For example, the policymaker might constrain himself to policies that take on a convenient parametric form. In contrast, strategic regularization is used to simplify the learning problem of the agent. The goal is to make the agent's behavior more predictable.

To formalize this, consider the policymaker's utility given a sample size n, a policy p that comes from an (ϵ, δ)-private strategy, and a true distribution π^s. I evaluate the policymaker's utility with respect to a worst-case mixed response π^r that satisfies the agent's regret bound (assumption 2).[6]

$$\begin{aligned} \mathrm{WC}_n(p, \epsilon, \delta, \pi^s) = & \min_{\pi^r} \mathrm{E}_{\pi^s, \pi^r}\left[u^P(p, r, s)\right] \\ & \text{s.t.} \quad \max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s, \pi^r}\left[u^A(p, r, s)\right] \le 4e^{\epsilon} \cdot \mathcal{RC}_n^A(p, \pi^s) + \delta \cdot \Delta^A(p) \end{aligned} \tag{4}$$

If the policymaker knew the true distribution π^s, he could safely ignore the sample and guarantee $(0, 0)$-privacy without loss of optimality. In that case, his worst-case optimal utility would be

$$\mathrm{SR}_n(\pi^s) = \max_p \mathrm{WC}_n(p, 0, 0, \pi^s) \tag{5}$$

The solution to (5) is called the *optimal strategically-regularized policy*.

In this formulation, unpredictable behavior by the agent tends to be costly for the policymaker. This reflects the worst-case optimization. As the agent's regret bound increases, the set of mixed responses π^r that satisfy this bound grows larger. For any given policy p, this gives the adversary more opportunities to lower the policymaker's utility, $\mathrm{E}_{\pi^s, \pi^r}\left[u^P(p, r, s)\right]$. On the other hand, I do not rule out unpredictability per se. The policymaker might find it advantageous to use a complicated policy p that leads to unpredictable behavior, as long as his worst-case utility $\mathrm{WC}_n(p, 0, 0, \pi^s)$ exceeds his worst-case utility under a less complicated policy p'.

Of course, the policymaker will typically not know the true distribution π^s in advance. Instead, he will learn about π^s using the sample. To that end, I construct an estimator $\hat{P}_n$ that uses the available data to approximate the optimal strategically-regularized policy.

4.1 Estimation

The estimator $\hat{P}_n$ is not especially complicated, but defining it will require a few steps. I start with a naive estimator and work my way from there, motivating each step in the process. Along the way, I introduce three tuning parameters $(\alpha, \beta, \epsilon)$ that do not vary in the sample size. I conservatively estimate the set of mixed responses that meet the agent's regret bound, and then guarantee sample privacy by adapting McSherry and Talwar's (2007) exponential mechanism. Finally, theorem 1 shows that the estimator $\hat{P}_n$ obtains the privacy guarantees that I assume in its definition.

[6] It is important to minimize over all mixed responses π^r. One may be tempted to minimize over pure responses r, but the resulting policy need not perform well. Here is the reason. Even if the agent uses a deterministic strategy, her response R_n still depends on the random sample. In most cases, the agent will observe a representative sample, and can choose a response that is optimal or nearly optimal. However, if the agent observes a very misleading sample, she may choose a very suboptimal response. From the policymaker's perspective, this can be represented as a mixed response. The agent makes a near-optimal response with high probability, and a very suboptimal response with low probability.

The naive estimator evaluates the policymaker's worst-case utility with respect to the empirical distribution $\hat{\pi}^s$. Formally,

$$\begin{aligned} \mathrm{WC}_n(p, \epsilon, \delta, \hat{\pi}^s) = &\min_{\pi^r} \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^P(p, r, s) \right] \\ &\text{s.t.} \quad \max_{r'} \mathrm{E}_{\hat{\pi}^s} \left[u^A \left(p, r', s \right) \right] - \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^A(p, r, s) \right] \leq 4e^{\epsilon} \cdot \mathcal{RC}_n^A(p, \hat{\pi}^s) + \delta \end{aligned} \tag{6}$$

For many problems in statistics, econometrics, and machine learning, replacing the true distribution with the empirical distribution is enough to come up with a good estimator. In this model, where the agent is learning in addition to the policymaker, three more changes are needed.

First, I add a buffer to the constraint in equation (6). Specifically,

$$\begin{aligned} &\min_{\pi^r} \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^P(p, r, s) \right] \\ \text{s.t.} \quad &\max_{r'} \mathrm{E}_{\hat{\pi}^s} \left[u^A \left(p, r', s \right) \right] - \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^A(p, r, s) \right] \leq (4e^{\epsilon} + 4) \cdot \mathcal{RC}_n^A(p, \hat{\pi}^s) + \delta \cdot \Delta^A(p) + \mathrm{BFR}_n \end{aligned} \tag{7}$$

The term BFR_n is defined as follows. Let $\alpha \in (0, 1)$ be a tuning parameter. Then

$$\mathrm{BFR}_n = 8\sqrt{\frac{2\ln 4}{n}} + 8\sqrt{-\frac{2\ln \exp(-n^{\alpha})}{n}}$$

Why is a buffer needed? It is because the constraint in equation (6) evaluates empirical regret, not true regret. That is, it evaluates the agent's utility with respect to the empirical distribution $\hat{\pi}^s$, not the true distribution π^s. If the sample is unrepresentative, mixed responses π^r that satisfy the regret bound may violate the constraint in equation (7) because their empirical regret overestimates their true regret. This would cause my estimator to incorrectly exclude mixed strategies that the agent might actually use. We can partially address this by adding a buffer to the regret bound, so that the empirical regret minus the buffer is unlikely to overestimate the true regret.

Lemma 1. *The buffer exceeds the difference between empirical and true regret, i.e.*

$$\begin{aligned} 4\mathcal{RC}_n^A(p, \pi^s) + \mathrm{BFR}_n \geq \Bigg| &\left(\max_{r'} \mathrm{E}_{\hat{\pi}^s} \left[u^A \left(p, r', s \right) \right] - \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^A(p, r, s) \right] \right) \\ &- \left(\max_{r'} \mathrm{E}_{\pi^s} \left[u^A \left(p, r', s \right) \right] - \mathrm{E}_{\pi^s, \pi^r} \left[u^A(p, r, s) \right] \right) \Bigg| \end{aligned}$$

for all mixed responses π^r and policies p, with probability no less than $1 - n_{\mathcal{P}} \exp(-n^{\alpha})$.

Second, I replace the Rademacher complexity $\mathcal{RC}_n^A(p, \pi^s)$ with an upper bound that does not depend on the distribution π^s. Specifically,

$$\begin{aligned} &\min_{\pi^r} \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^P(p, r, s) \right] \\ \text{s.t.} \quad &\max_{r'} \mathrm{E}_{\hat{\pi}^s} \left[u^A \left(p, r', s \right) \right] - \mathrm{E}_{\hat{\pi}^s, \pi^r} \left[u^A(p, r, s) \right] \leq (4e^{\epsilon} + 4) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta \cdot \Delta^A(p) + \mathrm{BFR}_n \end{aligned} \tag{8}$$

where $\overline{\mathcal{RC}}_n^A(p)$ is any distribution-free upper bound on the Rademacher complexity, i.e.

$$\overline{\mathcal{RC}}_n^A(p) \geq \max_{\pi^s} \mathcal{RC}_n^A(p, \pi^s) \tag{9}$$

There are plenty of distribution-free upper bounds on the Rademacher complexity, based on measures like the VC dimension, Pollard's pseudo-dimension, or the covering number. If all else fails, the fact that $\mathcal{R}$ is finite means I can obtain an upper bound using Massart's finite lemma. In addition, these distribution-free bounds are usually more tractable and interpretable than Rademacher complexity. This makes it easier to identify key variables driving policy complexity.

These two steps ensure that, with probability $1 - \exp(-n^\alpha)$, no mixed responses π^r that satisfy the regret bound are incorrectly excluded from the constraint in equation (8). The first step provides a conservative estimate of the agent's regret. The second step provides a conservative estimate of the bound on the agent's regret, which depends on the Rademacher complexity.

Third, I introduce white noise in order to control the sample privacy of the estimator $\hat{P}_n$. However, I cannot immediately apply existing privacy guarantees in my setting. This is due to an inherent circularity. On the one hand, sample privacy is a property of the estimator. On the other hand, in order to define the estimator, I need to specify the privacy parameters (ϵ, δ) in the agent's regret bound. Eventually, theorem 1 will show that I can resolve this circularity.

More concretely, I ensure sample privacy by adapting the exponential mechanism proposed by McSherry and Talwar (2007). Let $\beta \in (0, \alpha/2)$ be another tuning parameter. I add noise from the Gumbel distribution into the policymaker's objective function, i.e.

$$\nu_n(p) \sim \textsc{Gumbel}\left(0, n^{-\beta}\right)$$

For any given $\epsilon > 0$, this estimator will satisfy (ϵ, δ_n)-privacy, where δ_n depends on tuning parameters α, β, ϵ and is decreasing exponentially in n. More precisely,

$$\delta_n = n_{\mathcal{P}} \exp\left(-\frac{\epsilon^2}{2K^2} \cdot n^{\alpha - 2\beta}\right) \tag{10}$$

where the constant K is defined by

$$K := \max_p \max\left\{\frac{2\Delta^P(p)^2}{8\sqrt{2}}, \Delta^P(p)\right\} \tag{11}$$

Incorporating this into equation (8) gives a noisy, conservative estimate of the worst-case utility.

$$\begin{aligned} \widehat{\mathrm{WC}}_n(p) = \min_{\pi^r} & \ \mathrm{E}_{\hat{\pi}^s, \pi^r}\left[u^P(p, r, s)\right] + \nu_n(p) \\ \text{s.t.} \quad & \max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\hat{\pi}^s, \pi^r}\left[u^A(p, r, s)\right] \\ & \leq (4e^\epsilon + 4) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) + \mathrm{BFR}_n \end{aligned} \tag{12}$$

Definition 4. *The estimator $\hat{P}_n$ of the optimal strategically-regularized policy solves*

$$\hat{P}_n \in \arg\max_p \widehat{\mathrm{WC}}_n(p) \tag{13}$$

To summarize, the estimator depends on three tuning parameters. The parameter α controls how quickly the buffer BFR_n vanishes as n grows. The parameter β controls how quickly the privacy-preserving noise vanishes as n grows. The parameter ϵ controls how the two dimensions of sample privacy are balanced; decreasing ϵ means increasing δ_n, and vice-versa.

The next theorem verifies that this estimator obtains the privacy guarantees that were assumed in its definition. It holds for any fixed $\epsilon > 0$, with δ_n defined according to equation (10).

Theorem 1. *The estimator $\hat{P}_n$ guarantees $\left(\epsilon, \delta_n\right)$-sample privacy.*

The proof of theorem 1 relies on four lemmas, including known results. The key challenge is that $\widehat{\mathrm{WC}}_n(\cdot)$ is not a friendly object. It is a constrained minimization problem where the empirical distribution enters into both the objective and the constraint. And it is rather abstract, because I have made few assumptions on the game between the policymaker and the agent.

The first step is to show that the objective $\widehat{\mathrm{WC}}_n(\cdot)$ falls within a distance t of its mean, with high probability. This will be used to establish that $\widehat{\mathrm{WC}}_n(\cdot)$ satisfies a sample privacy property, which immediately implies that $\hat{P}_n$ satisfies that property. Intuitively, if $\widehat{\mathrm{WC}}_n(\cdot)$ varies substantially with the sample, then a substantial amount of noise $v_n(\cdot)$ will be needed to ensure privacy. This first step will limit how much $\widehat{\mathrm{WC}}_n(\cdot)$ varies with the sample.

I rely on a concentration inequality due to McDiarmid (1989). It relies on a bounded differences property that I will substantiate later on in this proof.

Lemma 2 (McDiarmid 1989)**.** *Suppose that $\widehat{\mathrm{WC}}_n(p)$ has the bounded differences property, where changing the ith sample realization from s to s' will change its value by at most c. Formally,*

$$\widehat{\mathrm{WC}}_n\left(p \mid S_1, \dots, S_{i-1}, s, S_{i+1}, S_n\right) - \widehat{\mathrm{WC}}_n\left(p \mid S_1, \dots, S_{i-1}, s', S_{i+1}, S_n\right) \le c \tag{14}$$

Then the following concentration inequality holds. For any $t > 0$,

$$\Pr\left[\widehat{\mathrm{WC}}_n(p) - \mathrm{E}\left[\widehat{\mathrm{WC}}_n(p)\right] \ge t\right] \le \exp\left(-\frac{2t^2}{nc^2}\right)$$

where the probability and expectation are over the sampling process.

It follows from McDiarmid's inequality and the union bound that

$$\Pr\left[\exists p \in \mathcal{P}, \widehat{\mathrm{WC}}_n(p) - \mathrm{E}\left[\widehat{\mathrm{WC}}_n(p)\right] \ge t\right] \le n_{\mathcal{P}} \exp\left(-\frac{2t^2}{nc^2}\right)$$

where $n_{\mathcal{P}}$ is the number of policies in $\mathcal{P}$. This is enough to guarantee sample privacy. To show this, I rely on the same reasoning that McSherry and Talwar (2007) used to establish differential privacy of their exponential mechanism.

Lemma 3. *The estimator $\hat{P}_n$ satisfies (ϵ, δ)-sample privacy where*

$$\epsilon = 2tn^{\beta} \quad \text{and} \quad \delta = n_{\mathcal{P}} \exp\left(-\frac{2t^2}{nc^2}\right)$$

for any $t > 0$, where c ensures the bounded differences property (14).

To establish the bounded differences property, I rely on the robustness lemma of Camara et al. (2020). Keep in mind that changing a sample realization S_i affects not only the objective in $\widehat{\mathrm{WC}}_n(p)$, but also the constraint. Since the game between the policymaker and agent is largely arbitrary, the impact of tightening or relaxing the constraint can be difficult to capture. However, it is possible to derive a loose bound that does not depend at all on the underlying structure of the game. It only relies on the fact that the worst case is being evaluated with respect to mixed strategies.

To state the robustness lemma, I need additional notation. Consider the policymaker's worst-case utility when the agent's regret is bounded by a constant $B \geq 0$, i.e,

$$\begin{aligned} \mathrm{WC}(p, B, \pi^s) = &\min_{\pi^r} \mathrm{E}_{\pi^s,\pi^r}\left[u^P(p, r, s)\right] \\ \text{s.t.} \quad &\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s,\pi^r}\left[u^A(p, r, s)\right] \leq B \end{aligned} \tag{15}$$

Lemma 4 (Camara et al. 2020)**.** *The worst-case utility* $\mathrm{WC}(p, B, \pi^s)$ *decreases smoothly in the bound B. That is, for any constants $B' > B > 0$,*

$$\mathrm{WC}\left(p, B', \pi^s\right) \geq \mathrm{WC}(p, B, \pi^s) - \Delta^A(p)\left(\frac{B' - B}{B}\right)$$

I use this result to establish and quantify the bounded differences property, as follows.

Lemma 5. *The random variable $\widehat{\mathrm{WC}}_n(p)$ satisfies the bounded differences property as long as*

$$c \geq \Delta^A(p)\left(\frac{2\Delta^P(p) \cdot n^{-1}}{(4e^{\epsilon} + 4) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta + \mathrm{BFR}_n}\right) + \Delta^P(p) \cdot n^{-1}$$

Now, I can define the parameter c. Recall from lemma 3 that δ depends on c. To avoid a circular definition, it is better that c not depend on δ. Moreover, c should not depend on the particular policy p. By lemma 5 and simple inequalities, it is sufficient to set

$$c \geq \max_p \left(\Delta^A(p)\left(\frac{2\Delta^P(p) \cdot n^{-1}}{(4e^{\epsilon} + 4) \cdot \overline{\mathcal{RC}}_n^A(p) + \mathrm{BFR}_n}\right) + \Delta^P(p) \cdot n^{-1}\right)$$

Technically, I could define c as the right-hand side on this expression and be done. But for the sake of interpretability and reducing notation, I define c using another upper bound. That is, let

$$c := Kn^{-\frac{1+\alpha}{2}}$$

where the constant K was defined above (11).

The last step of the proof is to derive δ_n. Recall from lemma 3 that the estimator $\hat{P}_n$ satisfies (ϵ, δ)-sample privacy where, after plugging in the value of c,

$$\epsilon = 2tn^{\beta} \quad \text{and} \quad \delta = n_{\mathcal{P}} \exp\left(-\frac{2t^2}{K^2} \cdot n^{\alpha}\right)$$

Since this holds for any value $t > 0$, I can invert the equation $\epsilon = 2tn^{\beta}$ to find the value $t = \epsilon/(2n^{\beta})$ that keeps the parameter ϵ constant. Plugging this value of t into δ gives me δ_n, as defined in equation (10). This completes the proof of theorem 1.

Having defined an estimator, we can ask: is it any good? The next two subsections address this question from an asymptotic perspective.

4.2 Convergence

In the limit as the sample size grows, the policymaker's payoff under the estimator $\hat{P}_n$ converges to his optimal payoff in a model where the distribution is common knowledge. That is the conclusion of proposition 3. This result will refer to three quantities of interest.

1. The common knowledge benchmark, $\text{CK}(\pi^s)$, describes the policymaker's optimal payoff when the distribution π^s is common knowledge. This was defined in equation (1).

2. The strategically-regularized benchmark describes the policymaker's optimal payoff when he knows the distribution π^s but the agent is still learning. This was defined in equation (5).

3. The performance of my estimator $\hat{P}_n$, as a strategy for the policymaker, is given by

$$\text{E}_{\pi^s}\left[\text{WC}_n\left(\hat{P}_n, \epsilon_n, \delta_n, \pi^s\right)\right]$$

It is easy to see that my estimator's performance is at most the strategically-regularized benchmark, which is at most the common knowledge benchmark. Less trivially, proposition 3 says that they all coincide in the limit, as $n \to \infty$.

To prove this result, I need another regularity assumption. To state that assumption, in turn, I need to define Rademacher complexity from the policymaker's perspective.[7]

Definition 5. *The policymaker's Rademacher complexity is defined over all policy-response pairs, i.e.*

$$\mathcal{RC}_n^P(\pi^s) = \text{E}_{\pi^s}\left[\sup_{p,r} \frac{1}{n}\sum_{i=1}^{n} \sigma_i \cdot u^P(p, r, S_i)\right]$$

[7] I use this quantity to bound the generalization error of the estimator $\hat{P}_n$. More precisely, for any given policy p, the generalization error is the difference between the performance of policy p according to the empirical distribution $\hat{\pi}^s$ and its actual performance under the true distribution π^s. Note that this error will generally depend on the agent's response r. To account for this, the policymaker's Rademacher complexity takes a supremum over responses r in addition to policies p. This ensures that the bound on the generalization error holds regardless of what the agent's response is or how it is influenced by the data.

where $\sigma_1, \dots, \sigma_n \sim$ UNIFORM are i.i.d. Rademacher random variables.

Both the agent and policymaker's Rademacher complexity should be diminishing at the typical $\tilde{O}(n^{-1/2})$ rate as the sample size grows. As usual, the tilde in $\tilde{O}$ means "up to log factors".

Assumption 3. *These two conditions apply to the agent and policymaker's Rademacher complexity.*

1. *There exists a constant K^A such that*

$$\overline{\mathcal{RC}}_n^A(p) \leq K^A n^{-1/2} \log n$$

for all policies p and sample sizes n, where K^A does not depend on p or n.

2. *There exists a constant K^P such that*

$$\mathcal{RC}_n^P(\pi^s) \leq K^P n^{-1/2} \log n$$

for all distributions π^s and sample sizes n, where K^P does not depend on π^s or n.

This assumption is both easy to satisfy and stronger than necessary. It is easy to satisfy because it holds whenever the agent's response space is finite. It also holds whenever the agent's optimization problem has a finite VC dimension, or a finite pseudo-dimension. And the assumption is stronger than necessary because the $\tilde{O}(n^{-1/2})$ rate is not needed for proposition 3. I do make use of this rate for theorem 1, in order to give concrete rates of convergence for my estimator.

As claimed, the estimator $\hat{P}_n$'s performance converges to the common prior benchmark in the limit as $n \to \infty$. Sandwiched between these two quantities is the strategically-regularized benchmark, which also converges.

Proposition 3. *Both the estimator $\hat{P}_n$'s performance and the strategically-regularized benchmark converge to the common knowledge benchmark, as $n \to \infty$. That is,*

$$\lim_{n\to\infty} \mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] = \lim_{n\to\infty} \mathrm{SR}_n(\pi^s) = \mathrm{CK}(\pi^s)$$

To prove this, I need some way to characterize the estimator $\hat{P}_n$'s performance. I do this by substituting its actual performance, which is hard to describe, with its estimated performance $\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)$. The estimated performance is easier to work with because that is what $\hat{P}_n$ is maximizing. The next lemma shows that this substitution is justified.

Lemma 6. *The estimated performance $\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)$ determines a lower bound on the the actual performance of the estimator $\hat{P}_n$. More precisely,*

$$\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] \geq \widehat{\mathrm{WC}}_n\left(\hat{P}_n\right) - n_{\mathcal{P}}\sqrt{3} \cdot n^{-\beta} - 2\mathcal{RC}_n^P(\pi^s) - n_{\mathcal{P}} \exp(-n^\alpha) \cdot \max_p \Delta^P(p)$$

This lower bound reflects three observations. First, by construction, $\widehat{\mathrm{WC}}_n(p)$ erred on the side of being too conservative with respect to the agent's empirical regret bound. All else equal, this would mean that $\widehat{\mathrm{WC}}_n(p)$ should lower bound $\mathrm{WC}_n(p)$ with high probability. Second, $\widehat{\mathrm{WC}}_n(p)$ involves sampling noise. This can lead to generalization error, where the policymaker expects a policy to perform better than it does. I can bound the generalization error using the policymaker's Rademacher complexity. Finally, $\widehat{\mathrm{WC}}_n(p)$ involves privacy-preserving noise. By construction, this is vanishing as the sample size grows.

In light of this lemma, in order to prove proposition 3 it is enough to show that

$$\lim_{n\to\infty} \mathrm{E}_{\pi^s}\left[\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] = \mathrm{CK}(\pi^s)$$

This is true because the privacy-preserving noise is vanishing as $n \to \infty$, the policymaker's empirical utility is converging in probability to his expected utility according to the true distribution, and the regret bound is vanishing as $n \to \infty$. In particular, Berge's maximum theorem implies that the policymaker's worst-case utility is continuous with respect to the regret bound.

4.3 Rate of Convergence

Having established convergence, it is natural to ask how quickly the estimator's performance converges to the common knowledge benchmark.

At first glance, there is no good answer. The next proposition clarifies. There exist games where the convergence of the strategically-regularized benchmark is arbitrarily slow.

Proposition 4. *For any $\gamma > 0$, there exists a game where the strategically-regularized benchmark has an $n^{-\gamma}$ rate of convergence, at best. That is,*

$$\mathrm{SR}_n(\pi^s) = \mathrm{CK}(\pi^s) - \Omega\left(n^{-\gamma}\right)$$

Furthermore, this game satisfies all of my regularity assumptions.

This result appears to be quite pessimistic, but it is not all that surprising. There happen to be games where the policymaker cannot guarantee the ideal outcome unless the agent has very precise distributional knowledge. In these cases, slow rates of convergence are inevitable. But that speaks to the fundamentals of the game itself, rather than to the quality of my estimator.

A more instructive question is to ask how the estimator $\hat{P}_n$'s rate of convergence compares to strategically-regularized benchmark. This is what I do in theorem 2. To that end, I assume that the strategically-regularized benchmark converges at an $O(n^{-\gamma})$ rate, where γ is arbitrary.

Assumption 4. *The strategically-regularized benchmark $\mathrm{SR}_n(\pi^s)$ converges to the common knowledge benchmark $\mathrm{CK}(\pi^s)$ at the rate $n^{-\gamma}$ for some $\gamma > 0$. That is,*

$$\mathrm{SR}_n(\pi^s) = \mathrm{CK}(\pi^s) - O(n^{-\gamma})$$

I also need an assumption that rules out trivial cases where the agent is indifferent between all of her responses. Roughly, I assume that for any distribution π^s and policy p, the agent's best response r is strictly better than her worst response r'.

Assumption 5. *For every distribution $\tilde{\pi}^s$ and policy p, there exist responses r, r' such that*

$$\mathrm{E}_{\tilde{\pi}^s}\left[\left(u^A(p,r,s) - u^A\left(p,r',s\right)\right)^2\right] \geq C^2$$

for some constant $C > 0$ that does not depend on $\tilde{\pi}^s, p, r, r'$.

Theorem 2 says that my estimator's rate of convergence is approximately optimal. I characterize the rate of convergence in terms of tuning parameters and then show how to optimize them.

Theorem 2. *Fix parameters α, β, ϵ where $\alpha \in (0,1)$, $\beta \in (0, \alpha/2)$, and $\epsilon > 0$. The policymaker's payoff under $\hat{P}_n$ converges to the common knowledge benchmark at the rate $n^{-\min(\gamma(1-\alpha),\beta)}$, i.e.*

$$\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] = \mathrm{CK}(\pi^s) - O\left(n^{-\min(\gamma(1-\alpha),\beta)}\right)$$

Note that the parameter ϵ does not affect the rate of convergence.[8]

The next corollary describes the rate of convergence when the tuning parameters are optimized. Generally, my estimator will only approximate the optimal rate of convergence $n^{-\gamma}$, with the difference reflecting the cost of sample privacy. For example, suppose that the strategically-regularized benchmark converges at a typical $n^{-1/2}$ rate. Then my estimator's rate of convergence can be set arbitrarily close to $n^{-1/4}$.

Corollary 1. *For any $\varepsilon > 0$, there exist parameter values α, β such that*

$$\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] = \mathrm{CK}(\pi^s) - O\left(n^{\frac{\gamma}{1+2\gamma}-\varepsilon}\right)$$

Proof. Set $\beta = \gamma/(1+2\gamma)$ and let α be slightly larger than 2β. □

Keep in mind that theorem 2 is, first and foremost, a possibility result. It says there exists an estimator, namely $\hat{P}_n$, that achieves a particular rate of convergence. It does not claim that this rate of convergence is the best possible. It relies on explicit finite sample bounds, but does not evaluate whether they are tight enough to be useful. It only provides limited guidance on how to choose the tuning parameter α in finite samples, and no guidance on how to choose ϵ, which does not affect the rate of convergence. Answering these questions effectively would likely require putting more structure on the underlying game.

To prove theorem 2, let us recall the proof of proposition 3. There, I showed that the empirical regret bound in the definition of $\widehat{\mathrm{WC}}_n(p)$ (12) was vanishing as the sample size grew. Similarly, the regret bound in the definition of $\mathrm{WC}_n(p, 0, 0, \pi^s)$ is vanishing as the sample size grows. For a given

[8] All else equal, it is better if ϵ is small, but it does not need to be small in order for the results to hold. It only needs to be constant. The reason is that the empirical regret bound depends on ϵ through

$$e^{\epsilon} \cdot \overline{\mathcal{R}C}_n^A(p)$$

This term vanishes as n grows because the Rademacher complexity vanishes as n grows.

sample size n, these bounds will take on different values. In general, they will shrink at different rates. And one bound involves empirical regret while the other involves regret with respect to the true distribution.

Despite these differences, we need to compare $\widehat{\mathrm{WC}}_n(p)$ with $\mathrm{WC}_n(p, 0, 0, \pi^s)$ to prove a result. As in theorem 1, these objects are too abstract to characterize directly. However, I can compare one abstract object with another: $\widehat{\mathrm{WC}}_n(p)$ for sample size n with $\mathrm{WC}_m(p, 0, 0, \pi^s)$ for a smaller sample size m. The idea is that if m is sufficiently small compared to n, then the regret bound for $\mathrm{WC}_m(p, 0, 0, \pi^s)$ is more conservative than the empirical regret bound for $\widehat{\mathrm{WC}}_n(p)$, even though the latter would be much more conservative if $m = n$. After accounting for some other differences, I can show that $\widehat{\mathrm{WC}}_n(p)$ is comparable to $\mathrm{WC}_m(p, 0, 0, \pi^s)$. Since I know the rate of convergence for the latter in m, I can determine the rate of convergence for the former in n.

The next lemma formalizes this argument.

Lemma 7. *The estimated performance $\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)$ with sample size n is comparable to the strategically-regularized benchmark with sample size*

$$m = \Theta(n^{1-\alpha})$$

evaluated with respect to the true distribution. More precisely,

$$\mathrm{E}_{\pi^s}\left[\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] \geq \mathrm{SR}_m(\pi^s) - n_{\mathcal{P}}\sqrt{3}\cdot n^{-\beta} - 2\mathcal{R}C_n^P(\pi^s) - n_{\mathcal{P}}\exp(-n^\alpha)\cdot \max_p \Delta^P(p)$$

This result relies critically on another lemma, which may be of independent interest. It is a lower bound on Rademacher complexity that relies on assumption 5 and Khintchine's inequality.

Lemma 8. *For any policy p and distribution π^s, the Rademacher complexity is bounded below by*

$$\mathcal{R}C_n^A(p, \pi^s) \geq \frac{C}{2\sqrt{2n}} \tag{16}$$

where the constant C was defined in assumption 5.

Combining lemma 7 with lemma 6 from the previous subsection, we have

$$\begin{aligned}
\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n, \epsilon, \delta_n, \pi^s\right)\right] &\geq \mathrm{SR}_m(\pi^s) - O\left(n^{-\beta}\right) - O\left(n^{-1/2}\right) \\
&= \mathrm{CK}(\pi^s) - O\left(m^{-\gamma}\right) - O\left(n^{-\beta}\right) - O\left(n^{-1/2}\right) \\
&= \mathrm{CK}(\pi^s) - O\left(m^{-\gamma}\right) - O\left(n^{-\beta}\right) \\
&= \mathrm{CK}(\pi^s) - O\left(n^{-\gamma(1-\alpha)}\right) - O\left(n^{-\beta}\right) \\
&= \mathrm{CK}(\pi^s) - O\left(n^{-\min(\gamma(1-\alpha),\beta)}\right)
\end{aligned}$$

The second line follows from assumption 4. The third line removes lower-order terms. The fourth line plugs in the value $m = \Theta\left(n^{2(1-\alpha)}\right)$. The last line takes the maximum over the two rightmost terms, and completes the proof of theorem 2.

5 Illustrative Examples

I argue strategic regularization can lead to new insights in four examples. Specifically, I consider vaccine distribution, prescription drug approval, performance-based pay, and product bundling. These examples are not intended to be as general or realistic as possible. Instead, they are meant to convey a core insight that motivates the use of strategic regularization in similar applications.

5.1 Vaccine Distribution

In a model of vaccine distribution, I show that strategic regularization can motivate a common practice: insisting on statistically-significant clinical trial results before delivering medical treatments. This kind of decision-making has been criticized (e.g. Wasserstein and Lazar 2016) and conflicts with the recommendations of the treatment choice literature (e.g. Manski 2019). Likewise, existing solutions do not support this practice in my model. But strategic regularization does. The intuition is that, when vaccine quality is not common knowledge, skepticism among the population can undermine a vaccine rollout. Since vaccine distribution involves fixed costs, it may be better to wait until clinical trial results are sufficiently persuasive before attempting to vaccinate the population.

Model. Consider a town of m agents that is afflicted by a disease. A new vaccine is being developed to treat this disease. However, in order to distribute this vaccine, the designer must invest in a treatment center at a fixed cost c. Given the treatment center, the designer can treat each agent at zero marginal cost. Therefore, the policymaker must decide whether to provide treatments ($T = 1$) at cost c, or not to treat ($T = 0$).

An agent's outcome Y depends on both whether she is treated, and whether she complies with the treatment. Let $C = 1$ indicate compliance, and $C = 0$ indicate noncompliance. Let Y_1 denote her outcome conditional on being successfully treated and let Y_0 denote her outcome otherwise. For simplicity, I assume that the agent has no private information about her outcome, so that compliance C is independent of the outcomes Y_0 and Y_1.

It remains to specify payoffs and the dataset. The agent tries to maximize her expected outcome:

$$\mathrm{E}[Y \mid C, T] = \mathrm{E}\big[Y_0 + C \cdot T \cdot (Y_1 - Y_0) \mid C, T\big] = \omega_0 + \omega_1 \cdot C \cdot T$$

The parameter ω_1 is called the average treatment effect (ATE). The designer wants to maximize the expected welfare minus costs, i.e.

$$m\mathrm{E}[Y \mid C, T] - c$$

Both the designer and the agents have access to clinical trial data where compliance is guaranteed. This includes n treated outcomes Y_1^i and n untreated outcomes Y_0^i. The key summary statistic is the sample average treatment effect, i.e.

$$\hat{\omega}_1 = \frac{1}{n}\sum_{i=1}^{n} Y_1^i - \frac{1}{n}\sum_{i=1}^{n} Y_0^i$$

This is a sufficient statistic for both the policymaker and the agent to optimize.

Existing Solutions. In the introduction, I alluded to several existing techniques for solving the policymaker's problem. In order to understand how strategic regularization generates new insights, it is useful to first look at the existing solutions.

If the outcome distribution is common knowledge, the policymaker treats iff

$$\omega_1 \geq \frac{c}{m}$$

That is, the ATE must exceed the per-capita cost. The reason is that the agent complies with the treatment iff $\omega_1 \geq 0$, which holds whenever the policymaker is inclined to treat.

The rational expectations assumption can motivate a qualitatively similar solution. The agent's behavior is unchanged, but now the policymaker faces a treatment choice problem. A natural strategy is to treat if and only if it is optimal according to the empirical distribution, i.e.

$$\hat{\omega}_1 \geq \frac{c}{m} \tag{17}$$

That is, the sample ATE must exceed the per-capita cost. This approach – empirical welfare maximization – has been widely studied in the literature on treatment choice (e.g. Manski 2004b, Stoye 2009, Kitagawa and Tetenov 2018, Mbakop and Tabord-Meehan 2021).

There are also two robust approaches to consider: maxmin utility and minmax regret. I make no assumptions on the agent's behavior but assume the policymaker knows the distribution (in order to avoid entirely trivial solutions). First, the maxmin optimal policy never treats. Note that it is not optimal to treat even if $\omega_1 \geq 0$. In the worst case, the policymaker incurs a cost $c > 0$ and the agent does not comply anyways. Second, the minmax regret policy never treats if $\omega_1 \leq c/m$. Otherwise, it treats iff

$$\omega_1 \geq \frac{2c}{m}$$

The reason is as follows. If the policymaker does not treat, the maximum regret $m\omega_1 - c$ occurs when the agents would have complied with treatment. If he does treat the population, the maximum regret c occurs when the agents decide not to comply.

None of these solutions justify statistically-significant clinical trial results as a precondition for treatment. Treatment was based on either the ATE or the sample ATE, but not on sample size or the "strength" of the evidence. This is not surprising. It is generally difficult to justify hypothesis testing as a tool for decision-making (e.g. Wasserstein and Lazar 2016, Manski 2019).

Strategic Regularization. Strategic regularization will motivate a form of statistical significance because, in this model, a treatment is only successful if agents agree to comply with said treatment. When clinical trial results are not statististically significant, agents may be sufficiently uncertain that they decide not to comply, even if compliance is optimal under the true distribution.

To substantiate this intuition, I solve for a variant of the optimal strategically-regularized policy. I maintain the assumption that the policymaker knows the true distribution. But I replace the agent's Rademacher complexity with an upper bound, for tractability, as with the estimator in section 4.1. This keeps the exposition relatively simple but preserves the key intuition. To bound the Rademacher complexity, I assume that the ATE is bounded, i.e. $\omega_1 \in [\underline{\omega}_1, \bar{\omega}_1]$ where $\bar{\omega}_1 > c/m > \underline{\omega}_1$. By Massart's finite lemma,

$$\overline{\mathcal{RC}}_n^A(T = 0) = 0 \quad \text{and} \quad \overline{\mathcal{RC}}_n^A(T = 1) = \frac{(\bar{\omega}_1 - \underline{\omega}_1)\sqrt{2 \ln 2}}{\sqrt{n}}$$

are valid upper bounds on the agent's Rademacher complexity.

The optimal policy only treats if the evidence is sufficiently compelling that even the most pessimistic agents will comply. Formally, it treats iff

$$\omega_1 \geq \frac{c}{m} \quad \text{and} \quad \omega_1 \geq \frac{(\bar{\omega}_1 - \underline{\omega}_1)\sqrt{2 \ln 2}}{\sqrt{n}} \tag{18}$$

The second term can be understood as a critical value; as usual, it vanishes at the rate $O(n^{-1/2})$. Essentially, this policy insists that the sample ATE be statistically significant before the policymaker decides to treat, in addition to the previous requirement that the ATE exceed per-capita cost.

In that sense, strategic regularization justifies a form of statistical significance, even if the policymaker knows the true distribution. If he does not know the distribution, he can use an estimator like the one developed in section 4.1. But the essential intuition survives. Namely, if the policymaker attempts to use an empirically-optimal policy (17) when the sample ATE is small relative to the sample size, he risks incurring the fixed cost c without having any agents comply.

5.2 Prescription Drug Approval

In a model of prescription drug approval by a regulator, I show that strategic regularization restricts doctors' ability to prescribe drugs that have not been proven effective in clinical trials. The threshold for approval increases as more drugs are approved. This is similar to stepwise methods for multiple hypothesis testing (e.g. Holm 1979, Romano and Wolf 2005). In contrast, in models with a common prior or rational expectations, the optimal policy is to approve all drugs. Essentially, this delegates the decision to doctors, who are better informed than the regulator. In my model, however, doctors may prescribe ineffective drugs if the clinical trial returns a "false positive" where an ineffective drug appears to be effective by random chance. Limiting the number of drugs approved can reduce the risk of false positives, and provide better welfare guarantees.

Model. A population of patients is afflicted by a disease. There are m treatments available, as well as a placebo. As in the previous example (section 5.1), let $\omega_j \in [\underline{\omega}, \bar{\omega}]$ be the average treatment effect of treatment j. The placebo's treatment effect is zero, and this is common knowledge. In addition, patients incur a private cost $c_j \in [0, \bar{c}]$ from treatment j, where the maximum cost $\bar{c} > \bar{\omega}$

exceeds the maximum treatment effect. For example, this could represent the patient's copay for a prescription drug. The patients are nonstrategic and accept whatever treatment is offered.

There is a regulator who approves treatments and a doctor who prescribes them. Formally, the regulator specifies a set $\mathcal{A} \subseteq \{1, \dots, m\}$ of approved treatments. Then the doctor either prescribes a treatment $j \in \mathcal{A}$ to a given patient, or prescribes the placebo. Both participants have access to clinical trial data with sample size n, where $\hat{\omega}_j$ is the sample ATE of treatment j. Both participants want to maximize the patient's expected outcome minus costs, i.e. $\omega_j - c_j$ for the chosen treatment j. But the doctor has an informational advantage. She knows patient costs c_j at the time of treatment choice. The regulator, however, does not know patient costs at the time of treatment approval.

Existing Solutions. If the ATEs are common knowledge, the optimal policy is to approve all treatments. Then the doctor will choose the treatment j that maximizes outcome minus costs, i.e. $\omega_j - c_j$. The regulator cannot do better with any other policy. Moreover, if the regulator excludes a treatment j, it is always possible that treatment j was the only treatment with low costs, e.g. where $c_j = 0$ and $c_i = \bar{c}$ for $i \neq j$. In that case, the regulator may regret excluding treatment j. Regardless of how the regulator deals with the uncertainty in costs, he strictly prefers to approve all treatments as long as the cost variation is nontrivial. These same conclusions hold under the rational expectations assumption, except that the regulator has even more incentive to defer because the doctor knows the ATE and the regulator does not.

Robust approaches, which make no assumptions about the agents beliefs, tend towards the other extreme of approving no treatments. As in section 5.1, I assume that the regulator knows the true ATE, as a best-case scenario for the robust approaches.

For example, the maxmin utility policy for the regulator approves no treatments. To see this, suppose the approved set $\mathcal{A}$ is nonempty. The worst case outcome occurs when all treatments $j \in \mathcal{A}$ share the maximum cost, $c_j = \bar{c}$, and the doctor chooses the worst treatment $j \in \mathcal{A}$ because she believes it to be highly effective. The regulator's utility is

$$\min_{j \in \mathcal{A}} \omega_j - \bar{c}$$

This is always negative, since $\omega_j < \bar{c}$. It is better to not approve any treatments, since this at least guarantees non-negative utility for the regulator.

The minimax regret policy also approves no treatments. Here is the argument. Note that the doctor might choose the placebo regardless of which treatments are approved. This occurs when all treatments have zero cost and the doctor believes they have zero effectiveness. In particular, let j^* be the treatment with the highest ATE. Suppose the regulator deviates from approvals $\mathcal{A}$ to approvals $\mathcal{A}'$ that includes j^*. The doctor breaks her indifference in favor of the placebo when presented with $\mathcal{A}$, but breaks her indifference in favor of treatment j^* when presented with $\mathcal{A}'$. The regulator's regret is given by $\omega_{j^*} = \max_j \omega_j$. Therefore, his worst-case regret is bounded below by ω_{j^*}. At the same time, approving no treatments guarantees that the regulator's regret is bounded above by ω_{j^*}. Therefore, approving no treatments minimizes worst-case regret.

Strategic Regularization. Strategic regularization will motivate solutions that are less extreme and more consistent with practice, where the regulator approves some treatments but not all. The reason is that, when multiple treatments are approved, the doctor essentially faces a multiple testing problem. The more treatments are approved, the greater the chances that the empirically-optimal treatment will not be as good as the data suggests. In other words, there may be false positives. Approving too many treatments can cause doctors to prescribe based on these false positives.

To substantiate this intuition, I solve a variant of the regulator's optimization problem with strategic regularization. There are three features to keep in mind. First, the regulator knows the true ATEs. This assumption can be relaxed using the estimator developed in section 4.1. Second, the regulator does not know the costs associated with different treatments. For the sake of concreteness, I assume that he minimizes worst-case regret. Third, the agent's Rademacher complexity is bounded above by

$$\overline{\mathcal{RC}}_n^A(\mathcal{A}) = \frac{(\bar{\omega} - \underline{\omega})\sqrt{2 \ln |\mathcal{A}|}}{\sqrt{n}}$$

This follows from Massart's finite lemma. In particular, the larger the number $|\mathcal{A}|$ of approved treatments, the more suboptimal the doctor's prescriptions may be. On the other hand, if too few treatments are approved, the doctor may be unable to prescribe optimal treatments.

To derive the minmax regret policy, I refer to two regret terms. First, the regulator's regret from not approving treatment j will be ω_j in the worst case. This occurs when treatment j has zero cost and all other treatments have maximum cost. It follows that worst-case regret is at least $\max_{j \notin \mathcal{A}} \omega_j$. Second, the regulator's regret from approving a treatment j will be $\overline{\mathcal{RC}}_n^A(\mathcal{A})$ in the worst case. Suppose $\mathcal{A}$ is nonempty. This level of regret can be achieved by setting the cost of all treatments that the doctor does not choose to $\bar{c}$, and the cost of the treatment j that the doctor does choose to $c_j = \omega_j + \overline{\mathcal{RC}}_n^A(\mathcal{A})$. The doctor will be indifferent between treatment j and the placebo.

The minmax regret policy minimizes the larger of these two regret terms. It begins by ordering treatments according to their ATE. Let $\omega_{(k)}$ be the kth highest ATE. The regulator approves the best treatment iff

$$\omega_{(1)} \geq \frac{(\bar{\omega} - \underline{\omega})\sqrt{2}}{\sqrt{n}}$$

Similarly, the regulator approves the kth best treatment iff

$$\omega_{(k)} \geq \frac{(\bar{\omega} - \underline{\omega})\sqrt{2 \ln k}}{\sqrt{n}}$$

As in section 5.1, treatments are approved only if they reach a critical value. It is easier to approve treatments with large sample sizes, as the critical value decreases in the sample size according to $O(n^{-1/2})$. Moreover, the critical value is increasing in the number of treatments already approved. As I mentioned earlier in this section, this is similar to stepwise methods for multiple hypothesis testing. That is not surprising, because the motivation for limiting the number of approved drugs is precisely to ensure that doctors are not misled by false positives.

The estimator in section 4.2 can be used to approximate the performance of this policy without knowing the true ATEs. It has one additional attribute worth highlighting. Sample privacy means that, with some probability, the regulator is not selecting the k treatments whose sample ATE is highest. To understand why that is, recall example 1, which I used to motivate sample privacy in the first place. That is a special case of this drug prescription model. More generally, it is well-known that using data to determine the hypotheses one wants to test will threaten the validity of those tests. There are several ways to get around this. For example, the regulator could use his prior beliefs to fix the k most promising treatments that he wants to evaluate and approve. In contrast, my approach is data-driven, and uses sample privacy to control how aggressively the data is used.

5.3 Performance Pay

Third, I consider a model of performance pay. An employer incentivizes an employee to exert costly effort by paying wages contingent on observed performance. Here, strategic regularization caps and flattens the wage schedule. Wages are zero until they reach some threshold level of performance. At that point, they jump and remain flat. When the sample size is small, the maximum wage is relatively small and easy to obtain. In contrast, the common prior solution compensates the employee only when the maximum possible performance is obtained, and pays a very large wage. In my model, this does not work well. If historical data is limited, it may not be obvious to the employee that it is worth investing effort for a small chance of receiving the bonus. Flatter contracts may be less potent, but they provide clearer incentives.

Model. assume that wages are denoted in dollars, to ensure finiteness

there exists a dataset consisting of outcomes conditional on shirk and outcomes conditional on work

this could be data that the manager collected through experimentation

———

I focus on a simple agency problem between an employer and an employee. Suppose the employee chooses whether to exert effort or not. Then, the employer receives a reward whose conditional distribution depends on the employee's effort. For clarity, I further assume that the conditional distributions satisfy the monotone likelihood ratio (MLR) property. Finally, the employee is compensated according to a wage schedule announced in advance. Following Sappington (1983), the agent is risk-neutral but the contract must satisfy limited liability (no negative wages). The penalized contract uses simple bounds on Rademacher complexity to recommend a threshold contract: the agent is paid a fixed wage if the reward exceeds some threshold, not paid if the reward falls below it, and partially paid at the threshold.

In this section, I present the baseline model, where all relevant parameters are common knowledge. There is a male principal who wants to incentivize a female agent to take desirable actions. The timing of the game is as follows: first, the principal commits to a wage schedule or contract w; second, the agent takes a hidden action x; third, the principal observes a noisy signal $\hat{x}$ of that action and pays the agent $w(\hat{x})$ based on the signal.

In the classical principal-agent problem, there is a (male) principal who wants to encourage effort by a (female) agent. The timing of the game is as follows: (1) the principal commits to a wage schedule, (2) the agent takes a hidden action, (3) nature randomly chooses a reward, (4) the agent is paid based on the reward, (5) the game concludes.

Let $R = \left\{r_1, \dots, r_m\right\} \subseteq \mathbb{R}$ be a finite reward space, in increasing order. Let $\theta \in \Delta R$ indicate a generic distribution over the rewards. The rewards are determined, stochastically, by the agent's hidden action. Let $A = \{0, 1\}$ be a binary action space, where $a = 0$ indicates no effort and $a = 1$ indicates effort. Let θ_0 be the true reward distribution conditional on no effort. Let θ_1 be the reward distribution conditional on effort. Let $Y \sim \theta_0$ and $Z \sim \theta_1$ be random variables, with y indicating (possibly counterfactual) reward from no effort and z indicating (possibly counterfactual) reward from effort.

Before the agent acts, the principal commits to a wage function $w : R \to \mathbb{R}_+$ that satisfies limited liability (i.e. $w(r) \geq 0$ for all $r \in R$). Having observed his reward, he transfers $w(r)$ to the agent. Both the principal and agent are risk-neutral. Given transfers t, action a, costs c, and reward r, the agent's utility is $u = t - ac$ while the principal's utility is $v = r - t$. That is, the agent seeks to lower her costs of effort, the principal seeks to maximize his rewards, and they respectively want to increase/decrease the wages paid. Here, limited liability prevents the principal from simply "selling the farm" and fully internalizing the agent's positive externality of effort.

Note: what about the screening aspect; multiple contracts? Feels like that should be a separate paper; I don't need to address everything at once, right? Like it's probably more important that I'm ignoring adverse selection than that I'm ignoring the possibility of screening.

The following assumption is made for simplicity.

Assumption 6. *The likelihood ratio $l(r) = \theta_1(r)/\theta_0(r)$ is weakly increasing in r.*

Let $Y_1, \ldots, Y_n \sim \theta_0$ and $Z_1, \ldots, Z_n \sim \theta_1$ be i.i.d. samples of rewards. Let $\hat{\theta}_0, \hat{\theta}_1 \in \Delta R$ denote the empirical distributions given by these samples. Given wage function w, let

$$\hat{a}_n = \mathbf{1}\left(\frac{1}{n}\sum_{i=1}^{n} w(Z_i) > c + \frac{1}{n}\sum_{i=1}^{n} w(Y_i)\right)$$

describe the agent's optimal action against the empirical distributions.

I will refer to an i.i.d. sample of n individuals, where individual i takes action X_i and generates a signal $\hat{X}_i$ with error E_i. The dataset consists of the signal errors, i.e.

$$S_n = \left(E_1, \ldots, E_n\right)$$

This dataset may be available in practice if, for example, an employer can directly observe performance X_i for a limited number of employees, but prefers to use a cheaper performance measure $\hat{X}$ for the remaining employees. However, this is obviously too restrictive: the whole motivation for models of moral hazard is that performance X_i is usually not directly observable.

Existing Solutions. Beyond engineering applications, the finite sample analysis aspires to formalize insights about market phenomena that cannot be expressed in the fully robust case ($n = 0$), the common prior case ($n = \infty$), or the prior-independent case ($n = \infty$ for the agent, $n < \infty$ for the principal). It succeeds in the principal-agent problem (with binary effort, risk-neutrality, and limited-liability). In the fully robust model, there is no contract that guarantees effort by the agent. The common prior model makes a seemingly extreme prediction: pay the agent if and only if the reward is the highest possible. In the finite sample setting, the principal encourages effort by pooling wages across a larger set of rewards.[9] The contrast with the prior-independent case will be discussed later.

It also contributes to a recent literature in robust contract design (e.g. Carroll 2015; Carroll and Meng 2016a,b; Dütting et al. 2019) that, in turn, builds on previous attempts to explain the ubiquity of simple contracts (e.g. Holmstrom and Milgrom 1987, 1991). In particular, Holmstrom and Milgrom (1991) motivate fixed wages when the principal is unable to measure some dimensions of the agent's performance. The penalized contract developed here suggests that learnability may be another motivation.

Valenzuela-Stookey (2020) "I have a paper on complexity in which I also discuss a principal-agent application (I presented this last year in the bag lunch). My approach is very different, but there is at least a superficial connection. In Section 3.5 I explore a model of learning from data, which I use to motivate my notion of complexity. This section is short and fairly self-contained. The application to a standard principal-agent problem is Section 7. I focus mainly on something like a worst-case approach ("cautious preferences"). I also get that step functions are optimal. I don't thing there's anything more than a superficial connection, but given that I also discuss learning from data and principal-agent problems I thought it worth pointing out. Perhaps there is some deeper connection between the learning model I discuss and the class of optimal contracts that you identify."

——

Common knowledge should be to pay zero except for the one outcome where payment is large; under MLR assumption, that one outcome is the highest possible

RE...agent knows distribution... designer is learning. Empirical not ideal because canâĂŹt assume the agent has the empirical beliefs. Probably want to add that buffer with the Lipschitz constant.

———

The optimal effort-inducing contract solves

$$w^* = \arg\max_{w} \mathrm{E}_{\theta_1}[Z - w(Z)] \quad \text{subject to} \quad \mathrm{E}_{\theta_0,\theta_1}[w(Z) - w(Y)] \geq c$$

[9] This pooling seems consistent with real-world experience, but can also be motivated as a response to risk aversion. Here is a case where the two explanations make different predictions. Consider a contractor who takes on a large number of heterogeneous jobs, for which she is individually rewarded (e.g. a large, well-diversified consulting firm). Barring confounders that cause correlation across jobs, her accumulated risk will be arbitrarily low, by the law of large numbers, so risk aversion would not play a role. Yet there may still be substantial ambiguity since the jobs are heterogeneous and not randomized. For that reason, we might still predict payment when performance is good but not ideal.

Throughout, I will take for granted that the principal weakly prefers to induce effort. Otherwise, the optimal contract is $w(r) = 0$ for all $r \in R$.

Proposition 5. *If θ_0, θ_1 are common knowledge, then*

$$w^*(r_i) = \begin{cases} \frac{c}{\theta_1(r_i) - \theta_0(r_i)} & i = m \\ 0 & i < m \end{cases}$$

That is, the agent is paid if and only if the realized reward is the highest possible.

Maximin utility: zero payments. Always possible that the agent decides not to work. If the agent does not work, best to not pay anything.

Minimax regret: probably also zero payments. For any contract, it is possible that the agent would have worked with just slightly higher incentives, and it is possible that the agent would not have worked anyways.

Strategic Regularization. Note: this might be wrong because the agent can use mixed strategies

Rademacher bound on agent is just Massart, depends on difference between maximum and minimum wage derive $\overline{\mathcal{RC}}_n^A(p)$ (this gives us an estimator)

Bound on policymaker Rademacher... you choose a continuous wage for m different outcomes... so a vector $\in \mathbb{R}_+^m$... presumably with an upper bound? derive $\overline{\mathcal{RC}}_n^P$ (this fives us a rate of convergence)

Comparative static... want to show that minimum wage decreases, maximum wage increases; under MLR assumption, threshold (of optimal strategically regularized policy) is increasing

The agent's expected utility (with respect to the true distribution) of action a is

$$U(a \mid \theta_0, \theta_1) = a\mathrm{E}_{\theta_1}[w(r) - c] + (1-a)\mathrm{E}_{\theta_0}[w(r)]$$

whereas the expected utility of the optimal action is

$$U^*(\theta_0, \theta_1) = \max_{a \in \{0,1\}} U(a \mid \theta_0, \theta_1)$$

Now we can solve for the maxmin-optimal contract. With high probability, this convinces the agent to exert effort by letting the optimality of effort exceed the agent's maximum suboptimality.

Definition 6. *The penalized contract solves*

$$\begin{aligned} w^{**} = &\arg\max_{w,\delta} \delta\mathrm{E}_{\theta_0}[Y - w(Y)] + (1-\delta)\mathrm{E}_{\theta_1}[Z - w(Z)] \\ &\text{subject to} \quad \mathrm{E}_{\theta_0,\theta_1}[w(Z) - w(Y)] \geq c + \kappa_1(\bar{w}, n, \delta) \end{aligned}$$

where $\bar{w} = \max_r w(r)$.

Proposition 6. *The penalized contract chooses $\bar{w}, \delta$ to optimize across the following class of threshold contracts. Fix $\bar{w}, \delta$ and define*

$$\alpha_j = \frac{c + \bar{w}\left(4\sqrt{\frac{2\log 2}{n}} - \sum_{i=j+1}^{m}\left(\theta_1(i) - \theta_0(i)\right)\right)}{\theta_1(j) - \theta_0(j)} + \sqrt{\frac{2\log(2/\delta)}{n}}$$

Let k be the largest j such that $l(r_j) > 1$ and $\alpha_j \leq \bar{w}$. If no such integer exists, set $w(r) = 0$ for all $r \in R$. Otherwise, set $w(r_i)$ to equal $\bar{w}$ if $i > k$, α_k if $i = k$, and 0 if $i < k$.

The penalized contract w^{**} builds on the Bayesian contract w^* by identifying the maximum wage $\bar{w}$ as an hindrance to learnability. By paying only if the highest reward r_m is realized, the Bayesian contract required (1) high wages in order to (2) exactly satisfy the agent's incentive constraint. It is not surprising that (2) is problematic, and indeed the penalized mechanism adds a buffer to the incentive constraints. What may have been less obvious is that (1), a consequence of concentrating on a single reward, is also problematic. When wages after any given reward are very high, small changes in the perceived probability of said reward can have a large impact on the perceived utility of effort. Therefore, a principal that relies on higher wages must allow for a larger buffer.

The penalized contract w^{**} builds on the Bayesian contract w^{*} by identifying the maximum wage $\bar{w}$ as an hindrance to learnability. When wages after any given reward are very high, small changes in the perceived probability of said reward can have a large impact on the perceived utility of effort. Therefore, a principal that relies on higher wages must allow for a larger buffer.

We can use comparative statics in n to better understand how our predictions may change as agents become better informed.

Corollary 2. *The maximum wage $\bar{w}$ and the threshold k of w^{**} are increasing in n. Moreover, there exist distributions θ_0, θ_1 where $w^{*} \neq w^{**}$ but both are nonzero.*

5.4 Product Bundling

Finally, I consider a model of product bundling. A firm has several products for sale and wants to sell them in a way that maximizes expected profit. Here, strategic regularization favors selling large bundles of products, or even bundling all products together into a grand bundle. This contrasts with prior work that suggests selling all items separately is robustly optimal, when there is not much information about consumer demand (Carroll 2017). In my model, the reason for bundling is that consumers learn about their value for the product through reviews. If there are many products, but few reviews per product, consumers can be confident in the value of the grand bundle while being uncertain about the value of any given product. In that case, all else equal, it is easier to convince consumers to buy the bundle.

Model. Consumers know their percentile for each product. They do not care about expected utility, but rather quantile utility. So if you knew the distribution, you would have no uncertainty.

marginals are known/learned; maxmin over correlation

assume that prices are denoted in dollars, to keep it finite

———

Consider every subset of $\{1, \ldots, m\}$ Assign a price to each subset... prices can be infinite must satisfy the property that if a menu can be constructed as the union of other menus, then the price of the original menu must be leq the price of the union of other menus

———

My assumption is that the WTP is jointly observed. This is a strong assumption. The designer may have access to WTP data, but may not be able to commit to releasing it to consumers. Meanwhile, both the designers and consumers will have access to public reviews of the product, but these will not necessarily identify the WTP. So, clearly, more work needs to be done in figuring out how consumers reach their assessments of product value. However, I argue that even this simpler form can highlight an important tradeoff when it comes to product bundling that is not apparent in either the common knowledge model or existing robust models (i.e. Gabe's multidimensional screening problem).

market research willingness to pay data – n sample points spread among m products, maybe random. We have data on values that we share with consumers. for each sample i and product j, you observe the value v_{ij}.

Existing Solutions. CK - want this to be Gabe Carroll's solution, i.e. offer each product separately (Carroll 2017)

RE - cite work on estimating the optimal price

maxmin - any pricing scheme that covers costs - reason being that in the worst case all agents believe they have zero value for the good

Strategic Regularization. penalty for size of menu offered

Proposition 7. *bound on EUM for quantile utility*

the bound will depend on the quantile... I guess you could maybe integrate over quantiles

This is also useful because it demonstrates that we don't need to use Rademacher complexity, or even usual notions of expected utility, to apply strategic regularization

Assumption 7. *regret is bounded by the term in 7*

let M be menu offered consumers automatically have the option to buy any subset. So if you offer k items on a menu, their number of actions is 2^k

$$\overline{\mathcal{RC}}_n^A(M) = \frac{\Delta^A(M)\sqrt{2|M|\ln 2}}{\sqrt{n}}$$

Proposition 8. *comparative static on p as $n \to \infty$... my guess is that number of items offered increases as n increases*

Consumer Information, Product Quality, and Seller Reputation

———

additive separability of the agent's objective is important here; that's what makes the learning problem easier when goods are sold separately, right?

–

A mechanism is a menu of possibly randomized product bundles with prices attached

The consumer has to choose a bundle

Obvious bound is number of bundles available. Grand bundle is really easy.

Have to assume utility is additively separable across goods, and linear in quantity, or WTP data doesnâĂŹt make sense.

Class of all menus is so rich, maybe not possible to get meaningful bound?

If you buy separately, you should have something like d times complexity of single product problem, posted price.

If you buy at most one, this is simple binary. Use Massart.

I guess if we use simple bound based on size of the menu, we get a log term in front of it. 2^d choices gives complexity O(d).

Choices are subsets of goods to buy, in this case.

Another interpretation: consumers are reading reviews, eg by critics.

Note that the producer only learns marginal distribution of values for each product. Like in the Carroll paper.

Social choice. Choose the thing that maximizes empirical utility âĂŤ or something like that. Data there is preferences. Showed that complexity of high dimensional is bounded by sum of complexity of low dimensional.

But that was when we had a high dimensional choice, right? Not when we had an arbitrary finite one.

Maybe just the finite bound when weâĂŹre talking about unit demand for each good.

AlsoâĂę I think the unit demand story just makes more sense, given the WTP data. Hard to think of data that would tell you about diminishing returns.

And in that case, if we restrict attention to deterministic mechanisms then the largest number of menus must be 2^d, since thatâĂŹs how many unique subsets of goods can be provided, and given two bundles that give the same allocation, the consumer will always choose the one with the lower price.

And I think it makes sense to restrict attention to deterministic mechanisms. LetâĂŹs be serious.

So the designerâĂŹs problem is basically to assign prices to every subset of goods.

Null option always needs to be represented.

If prices are set to infinity then those bundles should not be counted in the bound.

Odd aspect is that adding extra bundle to grand bundle can make the consumer less likely to choose grand bundle over outside option.

–

So to summarize, I think the bound here will depend on the size of the menu (or more precisely, the logarithm of the size. So offering grand bundle has smallest penalty, while selling separately has a penalty of d.

In contract example, penalty is Lipschitz constant of wage function plus maximum wage.

Netflix sells a single movie to a representative agent

- $\omega \in \Omega = \{0, 1\}$ movie quality
- $\theta_0 \in \Theta = [0, 1]$ probability that viewing experience is pleasurable
- $\mu \in M \subseteq \Delta(\Theta)$ agent's prior
- $Y_1, \dots, Y_n \sim \textsc{Bernoulli}(\theta_0)$ i.i.d. movie reviews
- $(x, p) \in A = [0, 1]^2$ probability of receipt and payment
- $u(x, p) = x\theta_0 - p$ agent's objective expected utility
- $v(x, p) = p$ principal's utility – need to adjust to allow possibility of non-sale

Robust solution: $x_n = 1$, $p_n = \inf_{\mu \in M} \mathrm{E}_{\theta \sim \mu}\left[\theta \mid Y_1, \dots, Y_n\right]$

- For M with priors whose density is bounded away from zero, $p_n \to_p \theta_0$ as $n \to \infty$

Trivial problem, will generalize in the next slide

- Still, want to mention some high-level questions you could ask
- At what rate does revenue p_n increase as n increases? What features of this model determine this rate, which don't?
- Fix two prices p, q and assertion "p (or q) obtains more revenue than q (or p)"
  - What is the largest n^* where assertion is true for some $\mu \in M$?
  - Let n be a reasonable sample size for some application
  - When $n < n^*$, refinements of $\{p, q\}$ are in some sense beyond the resolution of our theory (c.f. significant digits in chemistry)
- Having fixed M, n acts as a metric, useful for thinking about (rate of) convergence to (known, common) prior

Netflix sells m movies to a representative agent

- $\omega \in \Omega = \{0, 1\}^m$ movie quality
- $\theta_0 \in \Theta = [0, 1]^m$ probability that viewing experience is pleasurable (independent across movies)
- $\mu \in M \subseteq \Delta(\Theta)$ agent's prior
- $Y_{1j}, \dots, \dots, Y_{nj} \sim \textsc{Bernoulli}(\theta_{0j})$ i.i.d. movie reviews, independent across j

Consider two mechanisms. Separate sales.

- $(x, p) \in A = [0,1]^2$ probability of receipt and payment
- $u(x, p) = \sum_{j=1}^{m} (x_j \theta_{0j} - p_j)$ agent's objective expected utility
- $v(x, p) = \sum_{j=1}^{m} p_j$ principal's utility

Bundle

- $(x, q) \in A = [0,1]^2$ probability of receipt and payment
- $u(x, q) = \sum_{j=1}^{m} (x_j \theta_{0j}) - q$ agent's objective expected utility
- $v(x, q) = q$ principal's utility

Bundling has been seen as something that has value primarily when the distribution is known in great detail (**Carroll16**)

- Takes the view of a principal who lacks information

In this (super simple model)...

- Principal has no need to learn
- If μ were known, indifference: the robust solution to both would guarantee revenue $\inf_{\mu \in M} \sum_{j=1}^{m} \mathrm{E}_{\theta \sim \mu} [\theta_j \mid Y_1,$
- Likewise if $n = 0$, then indifference

If the agent is learning ($n > 0$) but their prior is uncertain (M large) is there value to bundling?

- I believe so, for technical reason that was already known (**Armstrong99**; **BB99**) but with a different interpretation

Observation 1. Let $m = 2^{12}$ be large and $n = 1$ be small

- Suppose μ expects average Netflix movie to be mediocre
- Suppose $Y_{1j} = 1$ for all j
- This is extremely unlikely unless θ_{0i} is nearly 1 for a large proportion of movies j
- MLE probably asserts all movies are high quality
- Contention: if movies are uniformly well-reviewed *and* you have confidence in the review process, you would conclude that the average Netflix movie is really high quality
    - As opposed to expecting average Netflix movie to be one good (but noisy) review above mediocre

  - **Manski04** might not, but that level of ambiguity aversion leads to extreme behavior (Stoye 2009)

- Restated: priors (distributions over Θ) will as a rule feature some amount of correlation across dimensions θ_i, θ_j

Observation 2. Let $m = 2^{12}$ be large and $n = 1$ be small

- The agent's decision of whether to buy the bundle is really simple: binary classification
  - Pooling reviews across all movies
  - Average error $\frac{1}{m} \sum_{j=1}^{m} \left(Y_{1j} - \theta_{0j} \right)$ will be small by LLN
- The agent's decision of optimal separate purchases is much less simple: m binary classifications
  - Small sample per movie means potentially high MSE
- Easy to convince a frequentist that it is worth buying Netflix at fair price; much harder to convince her to buy any individual movie
  - Contention is that Bayesian choice (for different priors) will be homogeneous for bundle purchase but heterogeneous for separate purchases
  - Heterogeneity means opportunity for adversary choosing μ

ntuition as follows:

- Define the "right" set of priors M
  - Requires more thought on how to translate frequentist intuition into a Bayesian restriction
- Let $m > 1$ be relatively large (many movies)
- Let $n > 0$ be relatively small (few reviews, or at least few trusted reviews, per movie)
- Consider (averaged) discount on the bundle price that we need to ensure all types purchase
- Should be less than the discount we need to ensure all types μ purchase an individual good
- If we don't discount prices on individual goods, adversary tries to choose priors s.t. valuation for many goods is just below the price and valuation for few goods is way above the price
- Micropayments for goods like podcasts may be less effective than subscription services because individuals find it harder to evaluate whether an individual series or episode is good or not than whether subscription plan is good or not, and so profits will be more susceptible to subjective beliefs

- Ofc toy model, not to be taken too seriously
  - For every Netflix there is an iTunes
  - For every Costco there is a Jewel-Osco
  - But highlights an important (?) consideration that seems to be obscured under existing frameworks

6 Related Literature

This work contributes to three research efforts. For robust mechanism design, it is a principled way to interpolate between two extremes: the common prior and prior-freeness. For learning in games, it provides a convenient behavioral assumption that does not rely on agents using a particular model or estimator. For data-driven mechanism design, it extends existing work to settings where the agent is learning, not just the policymaker. Below, I elaborate on each of these contributions.

Robust Mechanism Design. Robust mechanism design tries to relax the common prior assumption, as well as other knowledge assumptions used in mechanism design.

Initially, this literature focused on prior-free solution concepts that assumed no distributional knowledge whatsoever. Early on, Bergemann and Morris (2005) and Chung and Ely (2007) gave prior-free foundations for ex post incentive compatibility as a solution concept.[10] These papers worked with Harsanyi type spaces, where type profiles encode both the distribution of the state (or payoff type) as well as agents' higher-order beliefs. They sought to implement a social choice correspondence in any Bayes-Nash equilibrium of any type space.[11]

Prior-free solution concepts and the common prior assumption are two extreme cases, but there is a rich terrain that lies between them. For example, Oury and Tercieux (2012) propose *continuous implementation*. Given a type space that satisfies the common prior, the designer wants to implement a social choice correspondence in all type spaces that are arbitrarily close to the original one. Other researchers take a similar approach (e.g. Meyer-ter-Vehn and Morris 2011, Jehiel, Meyer-ter-Vehn, and Moldovanu 2012). Alternatively, Artemov et al. (2013) assume Δ-rationalizability (Battigalli and Siniscalchi 2003), where it is commonly known that the state distribution belongs to some pre-specified set Δ. In a similar spirit, Ollár and Penta (2017) assume that only pre-specified moments of the state distribution are common knowledge.

My work can also be seen as straddling the divide between prior-freeness on the one hand and the common prior on the other.[12] It is clearly inspired by the robustness literature, but its methods are

[10] Later contributions moved beyond ex post incentive compatibility (e.g. Börgers and Smith 2014, Börgers 2017).

[11] Prior-free approaches also developed in algorithmic game theory (Goldberg et al. 2006). This work focused on prior-free approximations to Bayesian-optimal mechanisms, whereas the economics literature focused on worst-case optimal mechanisms and characterizing which social choice correspondences were implementable.

[12] Granted, I avoid some of the issues related to strategic uncertainty that the prior literature has to deal with, due to my focus on single-agent mechanism design problems. But my learning-theoretic approach to relaxing the common prior assumption also seems to be compatible with multi-agent settings (c.f. Liang 2020).

different. Rather than specify a moment restriction or set Δ of plausible distributions, I assume that the agent has access to a dataset with sample size n. The parameter n controls how knowledgeable the policymaker and agent are supposed to be. It is a principled way to interpolate between prior-freeness ($n = 0$) and the common prior ($n = \infty$). This is true regardless of whether the dataset is interpreted literally (as in this paper), or as a stylized model of shared experience.

My model has three advantages relative to the nearest alternatives, namely Artemov et al. (2013) and Ollár and Penta (2017). First, it has few tuning parameters. The only parameter related to beliefs is the sample size n. Second, it makes it easier to decide "how much" robustness is required. I posit that researchers find it easier to gauge whether a sample size (or rate of convergence) is reasonable for their setting of interest, compared to an arbitrary set of beliefs or a set of moment restrictions. Third, it has a clear learning foundation. This is important because "robust" predictions can be quite sensitive to how one departs from the common prior assumption, so we need a good justification if we want to prioritize one over the other.[13]

Learning in Games. The literature on learning in games tries to replace prior knowledge or equilibrium assumptions with a more explicit process of learning from historical data. It is useful to divide this literature along two dimensions. First, whether data arises from repeated interaction (i.e. online learning) or random sampling (i.e. batch learning). Second, whether agents are learning about each other's strategies or about the state of nature. I am primarily concerned with models where agents learn about the state of nature through random sampling.

Liang (2020) is particularly relevant to my work. The author also studies incomplete information games where agents learn about the state through a finite dataset. In her model, agents adopt learning rules from a prespecified class of learning rules. If the learning rules are consistent, and converge uniformly, then predicted behavior is compatible with the common prior assumption in the limit as the sample size grows. In finite samples, predictions that hold under the common prior assumption (like the no-trade theorem) might not be necessarily true.

However, this paper differs from Liang (2020) in two respects. First, I commit to a particular class of learning rules: those that satisfy my regret bound. This class contains all learning rules that perform at least as well as empirical utility maximization, given the true distribution. By identifying a natural class of learning rules, I reduce the burden on researchers who want to use Liang's method. Second, my goal is policy design rather than predicting behavior. The new insights from my model do not come from agents learning per se. Instead, they come from the fact that policy choices can impact how quickly agents learn.

Researchers have also looked at the implications of statistical complexity for economic behavior. Some of this work considers the trade-off, from the agent's perspective, of choosing more or less complex statistical models to estimate (e.g. Al-Najjar and Pai 2014, Olea et al. 2021). Other work studies models of bounded rationality that can be motivated as a response to statistical complexity

[13] For example, if departures are defined using the product topology on the universal type space, then even small departures from the common prior can lead to drastic changes in predicted behavior (Lipman 2003; Rubinstein 1989; Weinstein and Yildiz 2007). In contrast, under the strategic topology, small departures from the common prior lead to small changes in predicted behavior (Chen et al. 2010; Dekel et al. 2006).

(e.g. Valenzuela-Stookey 2020, Jehiel 2005). In contrast, my work looks at how policy choices can make the agent's learning problem more or less complex. Furthermore, I do not assume that the agent is frequentist or that she relies on a particular statistical model.

Finally, researchers have also studied environments where agents' beliefs may not converge. This could be due to bounded rationality (Aragones et al. 2005; Haghtalab et al. 2021) or the fact that the environment is hopelessly complicated (Mailath and Samuelson 2020; Al-Najjar 2009). In particular, Al-Najjar (2009) relies on the notion of VC dimension, which is closely linked to the notion of Rademacher complexity used in this paper. Aside from this, however, the focus of these papers is different from my own. My assumptions imply that the agent can learn her optimal response to any policy, given sufficient data.

Data-driven Mechanism Design. The literature on data-driven mechanism design fuses robust mechanism design with learning in games. The goal is to combine microeconomic theory with data to provide more concrete policy recommendations. There is not much interaction between this literature and prior work in structural econometrics, presumably because it is driven by a community of computer scientists and microeconomic theorists, rather than empirical economists. As a result, the methods are somewhat different. The focus tends to be decision-theoretic, in line with Manski (2019), and there is less emphasis on estimating model parameters per se.

One prominent line of work studies the sample complexity of auction design. Here, the auctioneer lacks prior knowledge of the distribution of bidder values. Instead, he has access to a dataset, usually consisting of i.i.d. draws from the value distribution. A typical question is how many draws are needed in order for the auctioneer to guarantee near-optimal revenue with high probability (e.g. Balcan et al. 2008, Cole and Roughgarden 2014). Many of these papers rely on measures of learning complexity, like covering numbers (Balcan et al. 2008), pseudo-dimension (Morgenstern and Roughgarden 2015), and Rademacher complexity (Syrgkanis 2017). Gonçalves and Furtado (2020) use more familiar econometric methods towards a similar end.

These papers are focused on the auctioneer's learning problem, but ignore the bidders' learning problems. This is possible because of the application that they emphasize: auctions with dominant strategies, where agents have independent private values. In that context, there is no need for agents to learn about the value distribution. But this is not true in general. For example, in models with interdependent values, implementating reasonable outcomes in dominant strategies may be impossible (Jehiel, Meyer-ter-Vehn, Moldovanu, and Zame 2006). Alternatively, consider problems like monopoly regulation, contract design, or Bayesian persuasion. In these problems, the agent's optimal action depends on her beliefs over a hidden state of nature.

There is some prior work where both the policymaker and the agents are learning from data. For example, Camara et al. (2020) also study a single-agent policy design problem. Their model incorporates online learning, where data is generated over time through repeated interaction, and the data-generating process is arbitrary. In contrast, I consider batch learning, where data is generated from random sampling. Furthermore, Cummings et al. (2020) and Immorlica et al. (2020) consider agents that are learning from i.i.d. samples, respectively, in models of price discrimination and social learning. Both papers assume that agents' beliefs converge to the true distribution at a reasonable rate. In contrast, I assume that the agent's regret converges to zero at a reasonable rate. Although these assumptions should be mutually compatible, the advantage of regret bounds is that they make explicit how policies affect the complexity of the agent's learning problem.

7 Conclusion

In this paper, I proposed a modeling assumption that bypass a policymaker's lack of knowledge about the agent's beliefs: if the available data convincingly demonstrates some fact about the world, the agent should believe that fact. I studied this in the context of incomplete-information games where a policymaker commits to a policy, an agent responds, and both have access to a public dataset. I formalized the modeling assumption using concepts adapted from statistical learning theory, like regret, Rademacher complexity, and privacy. I showed that policies that are too complex in precise senses can be suboptimal because they lead to unpredictable behavior. To balance the benefits of policy complexity with its costs, I developed a method called strategic regularization and motivated it through both theoretical guarantees and illustrative examples.

The most important – and challenging – direction for future work is to turn strategic regularization towards real applications. One approach is to find highly-structured, data-rich settings where the statistical decision rules developed in this paper can actually be used. Particularly promising areas may lie in education or healthcare, where rich value-added measures have been used for policies like teacher compensation. Another approach is to treat strategic regularization as a desirable property even in the absence of an explicit dataset. Here, data would be seen as a metaphor for experiences that shape the agent's beliefs. Lab experiments could be used to determine whether more complex policies – as formalized in this paper – actually lead to suboptimal or unpredictable responses. In any case, extending this work to serious applications may require some theoretical extensions. For example, there may be settings where the distribution is only partially identified, where there are multiple agents interacting strategically, or where more efficient estimators can take advantage of particular problem structure. These are all worthwhile open questions.

References

Aragones, E., Gilboa, I., Postlewaite, A., & Schmeidler, D. (2005, December). Fact-free learning. *American Economic Review*, *95*(5), 1355–1368.

Artemov, G., Kunimoto, T., & Serrano, R. (2013). Robust virtual implementation: Toward a reinterpretation of the Wilson doctrine. *Journal of Economic Theory*, *148*(2), 424–447.

Balcan, M.-F., Blum, A., Hartline, J. D., & Mansour, Y. (2008). Reducing mechanism design to algorithm design via machine learning. *Journal of Computer and System Sciences*, *74*(8), 1245–1270.

Bartlett, P. L. & Mendelson, S. (2003, March). Rademacher and gaussian complexities: risk bounds and structural results. *J. Mach. Learn. Res. 3*, 463–482.

Battigalli, P. & Siniscalchi, M. (2003). Rationalization and Incomplete Information. *The B.E. Journal of Theoretical Economics*, *3*(1), 1–46.

Bergemann, D. & Morris, S. (2005). Robust mechanism design. *Econometrica*, *73*(6), 1771–1813.

Börgers, T. (2017, June). (no) foundations of dominant-strategy mechanisms: a comment on chung and ely (2007). *Review of Economic Design*, *21*(2), 73–82.

Börgers, T. & Smith, D. (2014, May). Robust mechanism design and dominant strategy voting rules. *Theoretical Economics*, *9*(2), 339–360.

Camara, M. K., Hartline, J. D., & Johnsen, A. (2020). Mechanisms for a no-regret agent: beyond the common prior. In *2020 ieee 61st annual symposium on foundations of computer science (focs)* (pp. 259–270).

Carroll, G. (2015, February). Robustness and linear contracts. *American Economic Review*, *105*(2), 536–63.

Carroll, G. (2017). Robustness and separation in multidimensional screening. *Econometrica*, *85*(2), 453–488.

Carroll, G. & Meng, D. (2016a). Locally robust contracts for moral hazard. *Journal of Mathematical Economics*, *62*, 36–51.

Carroll, G. & Meng, D. (2016b). Robust contracting with additive noise. *Journal of Economic Theory*, *166*, 586–604.

Chen, Y.-C., di Tillio, A., Faingold, E., & Xiong, S. (2010). Uniform topologies on types. *Theoretical Economics*, *5*(3), 445–478.

Chung, K.-S. & Ely, J. C. (2007). Foundations of dominant-strategy mechanisms. *The Review of Economic Studies*, *74*(2), 447–476.

Cole, R. & Roughgarden, T. (2014). The sample complexity of revenue maximization. In *Proceedings of the forty-sixth annual acm symposium on theory of computing* (pp. 243–252). STOC '14. New York, New York: ACM.

Cummings, R., Devanur, N. R., Huang, Z., & Wang, X. (2020). Algorithmic price discrimination. In *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '20. Salt Lake City, Utah, USA.

Dekel, E., Fudenberg, D., & Morris, S. (2006). Topologies on types. *Theoretical Economics*, *1*(3), 275–309.

Dütting, P., Roughgarden, T., & Talgam-Cohen, I. (2019). Simple versus optimal contracts. In *Proceedings of the 2019 acm conference on economics and computation* (pp. 369–387). EC '19. Phoenix, AZ, USA: ACM.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In S. Halevi & T. Rabin (Eds.), *Theory of cryptography* (pp. 265–284). Berlin, Heidelberg: Springer Berlin Heidelberg.

Goldberg, A. V., Hartline, J. D., Karlin, A. R., Saks, M., & Wright, A. (2006). Competitive auctions. *Games and Economic Behavior*, *55*(2), 242–269. Mini Special Issue: Electronic Market Design.

Gonçalves, D. & Furtado, B. (2020, August). *Statistical mechanism design: robust pricing and reliable projections*.

Haagerup, U. (1981). The best constants in the khintchine inequality. *Studia Mathematica*, *70*(3), 231–283.

Haghtalab, N., Jackson, M. O., & Procaccia, A. D. (2021). Belief polarization in a complex world: a learning theory perspective. *Proceedings of the National Academy of Sciences*, *118*(19).

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Holmstrom, B. & Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, *55*(2), 303–328.

Holmstrom, B. & Milgrom, P. (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, 24–52.

Immorlica, N., Mao, J., Slivkins, A., & Wu, Z. S. (2020). Incentivizing exploration with selective data disclosure. In *Proceedings of the 21st acm conference on economics and computation* (pp. 647–648). EC '20. Virtual Event, Hungary: Association for Computing Machinery.

Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory*, *123*(2), 81–104.

Jehiel, P., Meyer-ter-Vehn, M., & Moldovanu, B. (2012). Locally robust implementation and its limits. *Journal of Economic Theory*, *147*(6), 2439–2452.

Jehiel, P., Meyer-ter-Vehn, M., Moldovanu, B., & Zame, W. R. (2006). The limits of ex post implementation. *Econometrica*, *74*(3), 585–610.

Kitagawa, T. & Tetenov, A. (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, *86*(2), 591–616.

Liang, A. (2020, July). *Games of incomplete information played by statisticians*.

Lipman, B. L. (2003). Finite order implications of common priors. *Econometrica*, *71*(4), 1255–1267.

Mailath, G. J. & Samuelson, L. (2020, May). Learning under diverse world views: model-based inference. *American Economic Review*, *110*(5), 1464–1501.

Manski, C. F. (1993, January). Adolescent econometricians: how do youth infer the returns to schooling? In *Studies of supply and demand in higher education* (pp. 43–60). University of Chicago Press.

Manski, C. F. (2004a). Measuring expectations. *Econometrica*, *72*(5), 1329–1376.

Manski, C. F. (2004b). Statistical treatment rules for heterogeneous populations. *Econometrica*, *72*(4), 1221–1246.

Manski, C. F. (2019, December). *Econometrics for decision making: building foundations sketched by haavelmo and wald* (Working Paper No. 26596). National Bureau of Economic Research.

Mbakop, E. & Tabord-Meehan, M. (2021). Model selection for treatment choice: penalized welfare maximization. *Econometrica*, *89*(2), 825–848.

McDiarmid, C. (1989). On the method of bounded differences. In J. Siemons (Ed.), *Surveys in combinatorics, 1989: invited papers at the twelfth british combinatorial conference* (pp. 148–188). London Mathematical Society Lecture Note Series. Cambridge University Press.

McSherry, F. & Talwar, K. (2007). Mechanism design via differential privacy. In *48th annual ieee symposium on foundations of computer science (focs'07)* (pp. 94–103).

Meyer-ter-Vehn, M. & Morris, S. (2011). The robustness of robust implementation. *Journal of Economic Theory*, *146*(5), 2093–2104.

Morgenstern, J. & Roughgarden, T. (2015). The pseudo-dimension of near-optimal auctions. In *Proceedings of the 28th international conference on neural information processing systems - volume 1* (pp. 136–144). NIPS'15. Montreal, Canada: MIT Press.

Al-Najjar, N. I. (2009). Decision makers as statisticians: diversity, ambiguity, and learning. *Econometrica*, *77*(5), 1371–1401.

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

Olea, J. L. M., Ortoleva, P., Pai, M. M., & Prat, A. (2021, February). Competing models.

Ollár, M. & Penta, A. (2017, August). Full implementation and belief restrictions. *American Economic Review*, *107*(8), 2243–77.

Oury, M. & Tercieux, O. (2012). Continuous implementation. *Econometrica*, *80*(4), 1605–1637.

Romano, J. P. & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, *73*(4), 1237–1282.

Rubinstein, A. (1989). The electronic mail game: strategic behavior under "almost common knowledge". *The American Economic Review*, *79*(3), 385–391.

Sappington, D. (1983). Limited liability contracts between principal and agent. *Journal of Economic Theory*, *29*(1), 1–21.

Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics*, *151*(1), 7081.

Syrgkanis, V. (2017). A sample complexity measure with applications to learning optimal auctions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 5358–5365). NIPS'17. Long Beach, California, USA: Curran Associates Inc.

Valenzuela-Stookey, Q. (2020, September). *Subjective complexity under uncertainty*.

Wald, A. (1950). *Statistical decision functions*. Wiley: New York.

Wasserstein, R. L. & Lazar, N. A. (2016). The asa statement on p-values: context, process, and purpose. *The American Statistician*, *70*(2), 129–133.

Weinstein, J. & Yildiz, M. (2007). A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica*, *75*(2), 365–400.

A Omitted Proofs

A.1 Proof of Proposition 2

This proof will slightly more general than the proposition statement. Let $f : \mathcal{S}^n \rightarrow \mathbb{R}_+$ be an arbitrary function with upper bound $\bar{f}$. Suppose that there is an upper bound

$$\mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right)\right] \leq B$$

The goal is to find a similar upper bound on

$$\mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid P_n = p\right]$$

assuming that P_n is (ϵ, δ)-private. First, I use the privacy property to bound the following term.

$$\begin{aligned}
\mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid P_n = p, U\right] &= \sum_{S_1, \ldots, S_n} \mathrm{Pr}_{\pi^s}\left[S_1, \ldots, S_n \mid P_n = p, U\right] \cdot f\left(S_1, \ldots, S_n\right) \\
&= \sum_{S_1, \ldots, S_n} \frac{\mathrm{Pr}_{\pi^s}\left[S_1, \ldots, S_n \mid U\right] \cdot \mathrm{Pr}_{\pi^s}\left[P_n = p \mid S_1, \ldots, S_n\right]}{\mathrm{Pr}_{\pi^s}\left[P_n = p, U\right]} \cdot f\left(S_1, \ldots, S_n\right) \\
&\leq \sum_{S_1, \ldots, S_n} \frac{\mathrm{Pr}_{\pi^s}\left[S_1, \ldots, S_n \mid U\right] \cdot e^{\epsilon} \mathrm{Pr}_{\pi^s}\left[P_n = p, U\right]}{\mathrm{Pr}_{\pi^s}\left[P_n = p, U\right]} \cdot f\left(S_1, \ldots, S_n\right) \\
&= \sum_{S_1, \ldots, S_n} e^{\epsilon} \cdot \mathrm{Pr}_{\pi^s}\left[S_1, \ldots, S_n \mid U\right] \cdot f\left(S_1, \ldots, S_n\right) \\
&= e^{\epsilon} \cdot \mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid U\right]
\end{aligned}$$

Next, I use the upper bound to show that

$$\begin{aligned}
B &\geq \mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right)\right] \\
&= (1-\delta) \cdot \mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid U\right] + \delta \cdot \mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid \neg U\right] \\
&\geq (1-\delta) \cdot \mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid U\right] \\
&\geq (1-\delta) \cdot e^{-\epsilon} \cdot \mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid U, P_n = p\right] \\
&\geq (1-\delta) \cdot e^{-\epsilon} \cdot \mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid U, P_n = p\right] \\
&\quad + \delta \cdot e^{-\epsilon} \cdot \mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) - \bar{f} \mid \neg U, P_n = p\right] \\
&= e^{-\epsilon}\left(\mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid P_n = p\right] - \delta \cdot \bar{f}\right)
\end{aligned}$$

Finally, I rearrange the lower bound on B to obtain the desired result, i.e.

$$\mathrm{E}_{\pi^s}\left[f\left(S_1, \ldots, S_n\right) \mid P_n = p\right] \leq e^{\epsilon} \cdot B + \delta \cdot \bar{f}$$

A.2 Proof of Lemma 1

I want to show that

$$\begin{aligned}4\mathcal{R}C_n^A(p,\pi^s)+\mathrm{BFR}_n \geq \Bigg| &\left(\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p,r',s\right)\right]-\mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^A\left(p,r,s\right)\right]\right)\\ &-\left(\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right]-\mathrm{E}_{\pi^s,\pi^r}\left[u^A\left(p,r,s\right)\right]\right)\Bigg|\end{aligned}$$

for all mixed responses π^r and policies p, with probability no less than $1-n_{\mathcal{P}}\exp(-n^\alpha)$. For this purpose, it suffices to bound two quantities. First, observe that

$$\mathrm{E}_{\pi^s,\pi^r}\left[u^A\left(p,r,s\right)\right]-\mathrm{E}_{\hat{\pi}^s_n,\pi^r}\left[u^A\left(p,r,s\right)\right] \leq 2\overline{\mathcal{R}C}_n^A(p)+4\sqrt{\frac{2\ln(4/\kappa)}{n}} \tag{19}$$

with probability $1-\kappa$. This is the typical way of expressing regret bounds based on the Rademacher complexity (see e.g. Bartlett and Mendelson 2003). Second, observe that

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s_n}\left[u^A\left(p,r',s\right)\right]-\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right] \leq 2\overline{\mathcal{R}C}_n^A(p)+4\sqrt{\frac{2\ln(4/\kappa)}{n}} \tag{20}$$

This follows from the facts that

$$\mathrm{E}_{\hat{\pi}^s_n}\left[u^A\left(p,r^*,s\right)\right]-\mathrm{E}_{\pi^s}\left[u^A\left(p,r^*,s\right)\right] \leq 2\overline{\mathcal{R}C}_n^A(p)+4\sqrt{\frac{2\ln(4/\kappa)}{n}}$$

where $r^* \in \arg\max_r \mathrm{E}_{\hat{\pi}^s_n}\left[u^A\left(p,r,s\right)\right]$, and

$$\mathrm{E}_{\pi^s}\left[u^A\left(p,r^*,s\right)\right]-\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right] \leq 0$$

so

$$\mathrm{E}_{\hat{\pi}^s_n}\left[u^A\left(p,r^*,s\right)\right]-\mathrm{E}_{\pi^s}\left[u^A\left(p,r^*,s\right)\right]+\mathrm{E}_{\pi^s}\left[u^A\left(p,r^*,s\right)\right]-\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p,r',s\right)\right] \leq 2\overline{\mathcal{R}C}_n^A(p)+4\sqrt{\frac{2\ln(4/\kappa)}{n}}$$

Adding together the last two inequalities gives inequality (20). Adding together inequalities (19) and (20) gives the desired result. Applying the union bound across all policies p, the result holds with probability $1-n_{\mathcal{P}}\kappa$. Furthermore, the probability is uniform over all responses r, and therefore uniform across all mixed responses π^r. All that remains is to derive the probability κ and buffer BFR_n. Set $\kappa=\exp(-n^\alpha)$. Note that

$$\begin{aligned}4\sqrt{\frac{2\ln(4/\kappa)}{n}} &= 4\sqrt{\frac{2\ln(4\exp(-n^\alpha))}{n}}\\ &= 4\sqrt{\frac{2\ln 4-2\ln(\exp(-n^\alpha))}{n}}\end{aligned}$$

$$\begin{aligned}&\leq 4\sqrt{\frac{2\ln 4}{n}}+4\sqrt{-\frac{2\ln(\exp(-n^{\alpha}))}{n}}\\&\leq \mathrm{BFR}_n\end{aligned}$$

A.3 Proof of Lemma 3

Let $U \subseteq S^n$ be the set of all sample realizations $S_1, \dots, S_n$ where

$$\widehat{\mathrm{WC}}_n(p)-\mathrm{E}\left[\widehat{\mathrm{WC}}_n(p)\right]\leq t$$

By lemma 2, $\Pr_{\pi^s}[U]\geq 1-\delta$ where

$$\delta=\exp\left(-\frac{2t^2}{nc^2}\right)$$

To establish sample privacy, I need to show that

$$\Pr_{\pi^s}\left[\hat{P}_n=p \mid S_1,\dots,S_n\right]\leq e^{\epsilon}\cdot\Pr_{\pi^s}\left[\hat{P}_n=p,U\right]$$

for any sample realizations $\left(S_1,\dots,S_n\right)\in U$. Let the sample $\left(S_1',\dots,S_n'\right)\in U$ minimize

$$\Pr_{\pi^s}\left[\hat{P}_n=p \mid S_1',\dots,S_n'\right]$$

so that it suffices to show

$$\Pr_{\pi^s}\left[\hat{P}_n=p \mid S_1,\dots,S_n\right]\leq e^{\epsilon}\cdot\Pr_{\pi^s}\left[\hat{P}_n=p \mid S_1',\dots,S_n'\right] \tag{21}$$

I can characterize the distribution of $\hat{P}_n$ using standard results that link Gumbel error terms with exponential weights.

$$\Pr_{\pi^s}\left[\hat{P}_n=p \mid S_1,\dots,S_n\right]=\frac{\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p \mid S_1\dots,S_n\right)\right)}{\sum_{p'}\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p' \mid S_1\dots,S_n\right)\right)}$$

Using this, I can rewrite equation (21) and manipulate it as follows.

$$\begin{aligned}\frac{\Pr_{\pi^s}\left[\hat{P}_n=p \mid S_1,\dots,S_n\right]}{\Pr_{\pi^s}\left[\hat{P}_n=p \mid S_1',\dots,S_n'\right]}&=\frac{\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p \mid S_1\dots,S_n\right)\right)}{\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p \mid S_1'\dots,S_n'\right)\right)}\cdot\frac{\sum_{p'}\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p' \mid S_1'\dots,S_n'\right)\right)}{\sum_{p'}\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p' \mid S_1\dots,S_n\right)\right)}\\&\leq \exp\left(tn^{\beta}\right)\cdot\frac{\sum_{p'}\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p' \mid S_1'\dots,S_n'\right)\right)}{\sum_{p'}\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p' \mid S_1\dots,S_n\right)\right)}\end{aligned}$$

$$\le \exp\left(tn^{\beta}\right)\cdot\exp\left(tn^{\beta}\right)\cdot\frac{\sum_{p'}\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p'\mid S_1\ldots,S_n\right)\right)}{\sum_{p'}\exp\left(n^{\beta}\cdot\widehat{\mathrm{WC}}_n\left(p'\mid S_1\ldots,S_n\right)\right)}$$

$$\le \exp\left(2tn^{\beta}\right)$$

Therefore, $\hat{P}_n$ is (ϵ,δ)-private when $\epsilon = 2tn^{\beta}$.

A.4 Proof of Lemma 5

Recall the definition of $\mathrm{WC}(p, B, \pi^s)$ (15). This represents the policymaker's worst-case utility when the agent's regret is bounded by a constant $B \ge 0$. Note that

$$\widehat{\mathrm{WC}}_n(p) = \mathrm{WC}(p, B, \pi^s) \quad \text{where} \quad B = (4e^{\epsilon}+4)\cdot\overline{\mathcal{R}C}^{A}_{n}(p) + \delta + \mathrm{BFR}_n$$

Let $\hat{\pi}^s$ be the empirical distribution. Let $\tilde{\pi}^s$ be a modified empirical distribution where $S_i = s'$ instead of $S_i = s$. As I shift from $\hat{\pi}^s$ to $\tilde{\pi}^s$, the agent's empirical regret changes by at most $2\Delta^P(p)\cdot n^{-1}$. In particular, for any mixed response π^r,

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^A\left(p, r, s\right)\right] \le B$$

implies

$$\max_{r'} \mathrm{E}_{\tilde{\pi}^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\tilde{\pi}^s,\pi^r}\left[u^A\left(p, r, s\right)\right] \le B + 2\Delta^P(p)\cdot n^{-1}$$

Likewise, the policymaker's empirical utility changes by at most $\Delta^P(p)\cdot n^{-1}$. It follows from these two observations that

$$\widehat{\mathrm{WC}}_n(p \mid \hat{\pi}^s) \ge \mathrm{WC}\left(p, B + 2\Delta^P(p)\cdot n^{-1}, \tilde{\pi}^s\right) - \Delta^P(p)\cdot n^{-1}$$

where the notation $\widehat{\mathrm{WC}}_n(p \mid \hat{\pi}^s)$ is used to emphasize that $\widehat{\mathrm{WC}}_n(p)$ is being evaluated with respect to the empirical distribution $\hat{\pi}^s$. By the robustness lemma (4),

$$\mathrm{WC}\left(p, B + 2\Delta^P(p)\cdot n^{-1}, \tilde{\pi}^s\right) \ge \mathrm{WC}(p, B, \tilde{\pi}^s) - \Delta^A(p)\left(\frac{2\Delta^P(p)\cdot n^{-1}}{B}\right)$$

By definition, $\widehat{\mathrm{WC}}_n(p \mid \tilde{\pi}^s) = \mathrm{WC}(p, B, \tilde{\pi}^s)$. It follows that

$$\widehat{\mathrm{WC}}_n(p \mid \tilde{\pi}^s) - \widehat{\mathrm{WC}}_n(p \mid \hat{\pi}^s) \le \Delta^A(p)\left(\frac{2\Delta^P(p)\cdot n^{-1}}{B}\right) + \Delta^P(p)\cdot n^{-1}$$

Therefore, $\widehat{\mathrm{WC}}_n(p)$ satisfies the bounded differences property as long as

$$c \ge \Delta^A(p)\left(\frac{2\Delta^P(p)\cdot n^{-1}}{B}\right) + \Delta^P(p)\cdot n^{-1}$$

A.5 Proof of Lemma 6

I begin by making some observations and introducing some notation. Recall the empirical regret bound in the definition of $\widehat{\mathrm{WC}}_n(p)$ (12).

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^A\left(p, r, s\right)\right] \leq (4e^{\epsilon}+4)\cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) + \mathrm{BFR}_n \tag{22}$$

By lemma 1, any mixed response that satisfies this bound also satisfies

$$\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s,\pi^r}\left[u^A\left(p, r, s\right)\right] \leq 4e^{\epsilon}\cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p)$$

with probability $1 - n_{\mathcal{P}} \exp(-n^{\alpha})$, where the expectations are evaluated with respect to the true distribution π^s. This gives an upper bound for $\widehat{\mathrm{WC}}_n(p)$ with high probability, i.e.

$$\begin{aligned} f(p, \hat{\pi}^s) = &\min_{\pi^r} \mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^P\left(p, r, s\right)\right] + v_n(p) \\ &\text{s.t.} \quad \max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s,\pi^r}\left[u^A\left(p, r, s\right)\right] \leq 4e^{\epsilon}\cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) \end{aligned} \tag{23}$$

Let $\tilde{\pi}^r(p, \hat{\pi}^s)$ be the solution to the minimization problem (23). It is important to note that the set of feasible mixed responses no longer depends on the sample.

This proof consists of three parts. First, I want to bound the expected gap between $f(\hat{P}_n, \hat{\pi}^s)$ and $\widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)$. It follows from the preceding discussion that

$$\mathrm{E}_{\pi^s}\left[f(\hat{P}_n, \hat{\pi}^s) - \widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] \geq -n_{\mathcal{P}} \exp(-n^{\alpha}) \cdot \max_p \Delta^P(p)$$

Next, I want to bound the expected gap between $f(p, \hat{\pi}^s)$ and $f(p, \pi^s)$, i.e.

$$\begin{aligned} \mathrm{E}_{\pi^s}\left[\max_p \left(f(p, \hat{\pi}^s) - f(p, \pi^s)\right)\right] &= \mathrm{E}_{\pi^s}\left[\max_p \left(\mathrm{E}_{\hat{\pi}^s,\tilde{\pi}^r(p,\hat{\pi}^s)}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s,\tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right]\right)\right] \\ &\leq \mathrm{E}_{\pi^s}\left[\max_p \left(\mathrm{E}_{\hat{\pi}^s,\tilde{\pi}^r(p,\hat{\pi}^s)}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\hat{\pi}^s,\tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right]\right)\right] \\ &\quad + \mathrm{E}_{\pi^s}\left[\max_p \left(\mathrm{E}_{\hat{\pi}^s,\tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s,\tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right]\right)\right] \\ &\leq \mathrm{E}_{\pi^s}\left[\max_p \left(\mathrm{E}_{\hat{\pi}^s,\tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s,\tilde{\pi}^r(p,\pi^s)}\left[u^P\left(p, r, s\right)\right]\right)\right] \\ &\leq \mathrm{E}_{\pi^s}\left[\max_{p,\pi^r} \left(\mathrm{E}_{\hat{\pi}^s,\pi^r}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s,\pi^r}\left[u^P\left(p, r, s\right)\right]\right)\right] \\ &= \mathrm{E}_{\pi^s}\left[\max_{p,r} \left(\mathrm{E}_{\hat{\pi}^s}\left[u^P\left(p, r, s\right)\right] - \mathrm{E}_{\pi^s}\left[u^P\left(p, r, s\right)\right]\right)\right] \end{aligned}$$

At this point, it follows from the standard symmetrization argument that

$$\mathrm{E}_{\pi^s}\left[\max_p \left(f(p,\hat{\pi}^s) - f(p,\pi^s)\right)\right] \leq 2\mathcal{R}C_n^P(\pi^s)$$

Finally, I want to bound the expected gap between $\mathrm{WC}_n\left(p,\epsilon,\delta_n,\pi^s\right)$ and $f(\hat{P}_n,\pi^s)$. Note that $\mathrm{WC}_n\left(p,\epsilon,\delta_n,\pi^s\right) = f(p,\pi^s) - v_n(p)$. Furthermore, note that

$$\begin{aligned}
\mathrm{E}_{\pi^s}\left[f(\hat{P}_n,\pi^s) - \mathrm{WC}_n\left(\hat{P}_n,\epsilon,\delta_n,\pi^s\right)\right] &= \mathrm{E}_{\pi^s}\left[v_n(\hat{P}_n)\right] \\
&\leq \mathrm{E}\left[\max_p v_n(p)\right] \\
&\leq \mathrm{E}\left[\sum_p |v_n(p)|\right] \\
&= n_{\mathcal{P}}\mathrm{E}\left[|v_n(p)|\right] \\
&\leq n_{\mathcal{P}}\sqrt{\mathrm{E}\left[|v_n(p)|^2\right]} \\
&\leq n_{\mathcal{P}}\sqrt{\mathrm{E}\left[v_n(p)\right]^2 + \mathrm{Var}\left[v_n(p)\right]} \\
&\leq n_{\mathcal{P}}\sqrt{n^{-2\beta} + 2n^{-2\beta}} \\
&\leq n_{\mathcal{P}}\sqrt{3}\cdot n^{-\beta}
\end{aligned}$$

Combining these three steps gives us the desired result.

$$\begin{aligned}
&\mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n,\epsilon,\delta_n,\pi^s\right) - \widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] \\
&= \mathrm{E}_{\pi^s}\left[\mathrm{WC}_n\left(\hat{P}_n,\epsilon,\delta_n,\pi^s\right) - f(\hat{P}_n,\pi^s)\right] - \mathrm{E}_{\pi^s}\left[f(\hat{P}_n,\hat{\pi}^s) - f(\hat{P}_n,\pi^s)\right] + \mathrm{E}_{\pi^s}\left[f(\hat{P}_n,\hat{\pi}^s) - \widehat{\mathrm{WC}}_n\left(\hat{P}_n\right)\right] \\
&\geq -n_{\mathcal{P}}\sqrt{3}\cdot n^{-\beta} - \mathrm{E}_{\pi^s}\left[\max_p \left(f(p,\hat{\pi}^s) - f(p,\pi^s)\right)\right] - n_{\mathcal{P}}\exp(-n^{\alpha})\cdot\max_p \Delta^P(p) \\
&\geq -n_{\mathcal{P}}\sqrt{3}\cdot n^{-\beta} - 2\mathcal{R}C_n^P(\pi^s) - n_{\mathcal{P}}\exp(-n^{\alpha})\cdot\max_p \Delta^P(p)
\end{aligned}$$

A.6 Proof of Proposition 4

This example will exhibit a simple game where

$$\mathrm{SR}_n(\pi^s) = \mathrm{CK}(\pi^s) - \Omega\left(n^{-\gamma}\right)$$

The agent faces an estimation problem and cares about her accuracy. The policy space is a singleton; it is irrelevant. The state space $\mathcal{S} = [0,1]$ is the unit interval. The response space $\mathcal{R} = [0,1]$ is also the unit interval. The agent's response $r \in [0,1]$ is a prediction, subject to square loss, i.e.

$$u^A(p,r,s) = -(r-s)^2$$

The policymaker cares about the agent's accuracy with respect to a bliss point $s_0 \in [0, 1]$ that I will specify later. However, his sensitivity to inaccuracy is different from the agent, i.e.

$$u^P(p, r, s) = -|r - s_0|^{2\gamma}$$

I claim that there exists a distribution $\tilde{\pi}^s$ where the agent's regret bound is $\Omega(n^{-1})$. Let the bliss point $s_0 := \mathrm{E}_{\tilde{\pi}^s}[s]$ be the mean of s according to $\tilde{\pi}^s$. Let the distribution $\pi^s := \tilde{\pi}^s$. Existence follows from two observations. First, the mean square error of the maximum likelihood estimator is $O(n^{-1})$. Second, the maximum likelihood estimator is known be efficient.

To characterize the strategically-regularized benchmark, I need to consider responses that satisfy the agent's regret bound. One such response is is $r_n = \mathrm{E}_{\tilde{\pi}^s}[s] + \Omega(n^{-1/2})$. The policymaker's expected utility under r_n must be at least as large as the strategically-regularized benchmark, which is the worst case expected utility. That is,

$$\begin{aligned}
\mathrm{SR}_n(\pi^s) &\leq -|r_n - s_0|^{2\gamma} \\
&= -\left|\mathrm{E}_{\pi^s}[s] + \Omega(n^{-1/2}) - \mathrm{E}_{\tilde{\pi}^s}[s]\right|^{2\gamma} \\
&= -\left(\Omega(n^{-1/2})\right)^{2\gamma} \\
&= -\Omega(n^{-\gamma}) \qquad (24)
\end{aligned}$$

Next, consider the common knowledge benchmark. The agent will predict the mean, $r = \mathrm{E}_{\pi^s}[s]$, and the policymaker's expected utility will be

$$\mathrm{CK}(\pi^s) = \left|\mathrm{E}_{\pi^s}[s] - s_0\right|^{2\gamma} = \left|\mathrm{E}_{\tilde{\pi}^s}[s] - \mathrm{E}_{\tilde{\pi}^s}[s]\right|^{2\gamma} = 0 \qquad (25)$$

I can combine equations (24) and (25) to show

$$\mathrm{SR}_n(\pi^s) \leq \mathrm{CK}(\pi^s) - \Omega(n^{-\gamma})$$

This completes the first part of the proof.

Next, I need to verify that this game satisfies (in particular) assumption 3. To do this, I need to introduce the pseudodimension: a method for bounding the Rademacher complexity. The following definition is specialized to the agent's Rademacher complexity.

Definition 7. *A vector $(w_1, \ldots, w_n) \in \mathbb{R}^n$ is a* witness *for a vector $(S_1, \ldots, S_n)$ if, for any realizations $(\sigma_1, \ldots \sigma_n) \in \{-1, 1\}^n$, there exists a response r such that*

$$\operatorname{sign}\left(-\left(r - S_i\right)^2 - w_i\right) = \sigma_i \qquad (26)$$

A vector $(S_1, \ldots, S_n)$ is shattered *if it has a witness $(w_1, \ldots, w_n)$. The* pseudo-dimension *is the largest integer m such that some vector $(S_1, \ldots, S_m)$ is shattered.*

Claim 1. *The pseudo-dimension is at most 2.*

Since the pseudo-dimension is bounded, the agent's Rademacher complexity is $\tilde{O}(n^{-1/2})$.

Proof. For the sake of contradiction, suppose that the vector $S_1, \dots, S_n$ is shattered for $n > 2$. By condition (26), $\sigma_i = 1$ means that S_i is within some distance d_i of r, where d_i depends on w_i and γ. Define n intervals $I_1, \dots, I_n$ where $I_i = [S_i - d_i, S_i + d_i]$. Then $\sigma_i = 1$ means $r \in I_i$, and $\sigma_i = 0$ means $r \notin I_i$. Let $f(r)$ be the set of intervals I_i such that $r \in I_i$. Each vector σ corresponds to a unique element in the range of $f(r)$.

I claim that the range of $f(r)$ has at most $2n + 1$ elements. If we list the n left endpoints and the n right endpoints of intervals, in order, these define a different set of $2n + 1$ intervals. Within each interval J, we can move r from the left to the right, without entering or exiting any interval I_i. Therefore, f is invariant over each interval J. Since there are at most $2n + 1$ intervals J, the range of $f(r)$ must have at most $2n + 1$ elements.

However, this leads to a contradiction. There are 2^n distinct values of the vector σ. But each vector σ must correspond to a unique element in the range of r, and there are only $2n + 1$ such elements. When $n = 3$, $2^n = 8$ but $2n + 1 = 7$. When $n > 3$, the discrepancy is even larger. Therefore, the vector $S_1, \dots, S_n$ does not have a witness when $n > 3$. It follows from the definition that the pseudo-dimension is at most 2. □

Next, consider the policymaker's Rademacher complexity. Note that

$$\max_r \sum_{i=1}^{n} \sigma_i |r - s_0|^{2\gamma}$$

has only three possible solutions: $r = s_0$, $r = 0$, or $r = 1$. Without loss of generality, I can restrict the response space to $\{0, s_0, 1\}$. It follows from Massart's finite lemma that the policymaker's Rademacher complexity is $O(n^{-1/2})$.

A.7 Proof of Lemma 7

Recall the empirical regret bound in the definition of $\widehat{\mathrm{WC}}_n(p)$ (12).

$$\max_{r'} \mathrm{E}_{\hat{\pi}^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\hat{\pi}^s, \pi^r}\left[u^A\left(p, r, s\right)\right] \leq (4e^{\epsilon} + 4) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) + \mathrm{BFR}_n \tag{27}$$

I want to argue that every mixed response π^r that satisfies this empirical regret bound also satisfies the regret bound in the definition of $\mathrm{WC}_m(p, 0, 0, \pi^s)$, i.e.

$$\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s, \pi^r}\left[u^A\left(p, r, s\right)\right] \leq 4 \cdot \mathcal{RC}_m^A(p, \pi^s) \tag{28}$$

At least, this should hold with high probability. Let π^r be a mixed response satisfying the empirical regret bound (27). By lemma 1, with probability $1 - n_{\mathcal{P}} \exp(-n^{\alpha})$, we have

$$\max_{r'} \mathrm{E}_{\pi^s}\left[u^A\left(p, r', s\right)\right] - \mathrm{E}_{\pi^s, \pi^r}\left[u^A\left(p, r, s\right)\right] \leq (4e^{\epsilon} + 8) \cdot \overline{\mathcal{RC}}_n^A(p) + \delta_n \cdot \Delta^A(p) + 2\mathrm{BFR}_n \tag{29}$$

This puts the agent's regret in terms of the true distribution. Next, I claim that the right-hand side of inequality (29) is $\Theta(m^{-1/2})$. There are three terms to consider. The first term is $\tilde{O}(n^{-1/2})$ by assumption 3. The second term is decreasing exponentially in n, since $\alpha < 2\beta$. The third term is $\Theta(n^{(\alpha-1)/2})$, and it is leading since $\alpha > 0$. Plugging in the value of m gives us $\Theta(m^{-1/2})$. Finally, note that as long as there is a sufficiently large constant in front of m, we have

$$\begin{aligned}(4e^{\epsilon}+8)\cdot\overline{\mathcal{RC}}_n^A(p)+\delta_n\cdot\Delta^A(p)+2\mathrm{BFR}_n &\le 4\cdot\frac{C}{2\sqrt{2m}}\\ &\le 4\cdot\mathcal{RC}_m^A(p,\pi^s) \qquad (30)\end{aligned}$$

where the last line follows from lemma 8. Combining inequalities (29) and (30) gives us the desired inequality (28), with probability $1-n_{\mathcal{P}}\exp(-n^{\alpha})$.

I have established that the set of mixed responses that $\widehat{\mathrm{WC}}_n(p)$ minimizes over is, with high probability, a subset of the set of mixed responses that $\mathrm{WC}_m(p,0,0,\pi^s)$ minimizes over. All that remains is to compare the policymaker's objective under $\widehat{\mathrm{WC}}_n(p)$ with his objective under $\mathrm{WC}_m(p,0,0,\pi^s)$. This compares expected utility under the empirical distribution, plus privacy-preserving noise, to expected utility under the true distribution. But this is precisely the situation we found ourselves in during the proof of lemma 6. I can apply the same bounds here to complete the proof.

A.8 Proof of Lemma 8

Recall the definition of Rademacher complexity:

$$\begin{aligned}\mathcal{RC}_n^A(p,\pi^s) &= \frac{1}{n}\mathrm{E}_{\pi^s}\left[\max_r\sum_{i=1}^n\sigma_i\cdot u^A\left(p,r,S_i\right)\right]\\ &= \frac{1}{n}\mathrm{E}_{\pi^s}\left[\mathrm{E}_{\pi^s}\left[\max_r\sum_{i=1}^n\sigma_i\cdot u^A\left(p,r,S_i\right)\mid S_1,\dots,S_n\right]\right]\end{aligned}$$

where the second equality follows from the law of iterated expectations. To bound the Rademacher complexity, it suffices to bound the interior expectation. Observe that

$$\begin{aligned}&\mathrm{E}_{\pi^s}\left[\max_r\sum_{i=1}^n\sigma_i\cdot u^A\left(p,r,S_i\right)\mid S_1,\dots,S_n\right]\\ &=\max_{r'}\mathrm{E}_{\pi^s}\left[\max_r\left(\sum_{i=1}^n\sigma_i\cdot u^A\left(p,r,S_i\right)-\sum_{i=1}^n\sigma_i\cdot u^A\left(p,r',S_i\right)+\sum_{i=1}^n\sigma_i\cdot u^A\left(p,r',S_i\right)\right)\mid S_1,\dots,S_n\right]\\ &=\max_{r'}\mathrm{E}_{\pi^s}\left[\max_r\left(\sum_{i=1}^n\sigma_i\cdot\left(u^A\left(p,r,S_i\right)-u^A\left(p,r',S_i\right)\right)\right)+\sum_{i=1}^n\sigma_i\cdot u^A\left(p,r',S_i\right)\mid S_1,\dots,S_n\right]\\ &=\max_{r'}\mathrm{E}_{\pi^s}\left[\max_r\left(\sum_{i=1}^n\sigma_i\cdot\left(u^A\left(p,r,S_i\right)-u^A\left(p,r',S_i\right)\right)\right)\mid S_1,\dots,S_n\right] \qquad (31)\end{aligned}$$

$$= \max_{r'} \mathrm{E}_{\pi^s} \left[\max_{r} \left(\sum_{i=1}^{n} \sigma_i \cdot \left(u^A \left(p, r, S_i\right) - u^A \left(p, r', S_i\right)\right) \right)^+ \mid S_1, \ldots, S_n \right] \tag{32}$$

$$\geq \max_{r,r'} \mathrm{E}_{\pi^s} \left[\left(\sum_{i=1}^{n} \sigma_i \cdot \left(u^A \left(p, r, S_i\right) - u^A \left(p, r', S_i\right)\right) \right)^+ \mid S_1, \ldots, S_n \right] \tag{33}$$

$$= \frac{1}{2} \max_{r,r'} \mathrm{E}_{\pi^s} \left[\left| \sum_{i=1}^{n} \sigma_i \cdot \left(u^A \left(p, r, S_i\right) - u^A \left(p, r', S_i\right)\right) \right| \mid S_1, \ldots, S_n \right] \tag{34}$$

$$\geq \frac{1}{2\sqrt{2}} \max_{r,r'} \sqrt{\sum_{i=1}^{n} \left(u^A \left(p, r, S_i\right) - u^A \left(p, r', S_i\right)\right)^2} \tag{35}$$

$$\geq \frac{C\sqrt{n}}{2\sqrt{2}}$$

The first two equalities follow from algebraic manipulations. Line (31) follows from the fact that

$$\mathrm{E}_{\pi^s} \left[\sum_{i=1}^{n} \sigma_i \cdot u^A \left(p, r', S_i\right) \mid S_1, \ldots, S_n \right] = 0$$

Line (32) follows from the fact that setting $r = r'$ ensures that the interior sum is zero, so that the maximum over all r is non-negative. Line (33) follows from Jensen's inequality. Line (34) follows from the fact that the sum inside the expectation is symmetrically distributed around zero. To see this, let X be a symmetric random variable with mean zero. Then

$$\begin{aligned}
\mathrm{E}[|X|] &= \Pr[X = 0] \cdot 0 + \Pr[X > 0] \cdot \mathrm{E}[X \mid X \geq 0] + \Pr[X < 0] \cdot \mathrm{E}[-X \mid X < 0] \\
&= \Pr[X > 0] \cdot \mathrm{E}[X \mid X > 0] + \Pr[X > 0] \cdot \mathrm{E}[X \mid X > 0] \\
&= 2 \cdot \Pr[X > 0] \cdot \mathrm{E}[X \mid X > 0] \\
&= 2 \cdot \mathrm{E}\left[X^+\right]
\end{aligned}$$

Line (35) follows from Khintchine's inequality, with constants derived by Haagerup (1981). Finally, the last inequality follows from assumption 5.

A.9 Proof of Proposition 5

Assume binding constraint.

$$\frac{1}{n} \sum_{i=1}^{n} \theta_1(r_i) \left(r_i - w(r_i)\right)$$

$$\frac{1}{n} \sum_{i=1}^{n} \theta_1(r_i) \left(w(r_i) - c\right) = \frac{1}{n} \sum_{i=1}^{n} \theta_0(r_i) w(r_i)$$

$$\frac{1}{n}\sum_{i=1}^{n}\theta_1(r_i)r_i - \frac{1}{n}\sum_{i=1}^{n}\theta_1(r_i)w(r_i)$$

$$\frac{1}{n}\sum_{i=1}^{n}\theta_1(r_i)w(r_i) = c + \frac{1}{n}\sum_{i=1}^{n}\theta_0(r_i)w(r_i)$$

$$\frac{1}{n}\sum_{i=1}^{n}\theta_1(r_i)r_i - c - \frac{1}{n}\sum_{i=1}^{n}\theta_0(r_i)w(r_i)$$

$$\frac{1}{n}\sum_{i=1}^{n}\theta_0(r_i)w(r_i) \geq 0$$

Linear programming problem? Hence extreme solution

A.10 Proof of Proposition 6

$$\min \frac{1}{n}\sum_{i=1}^{n}\theta_1(r_i)w(r_i)$$

subject to

$$\frac{1}{n}\sum_{i=1}^{n}\theta_1(r_i)w(r_i) = c + \frac{1}{n}\sum_{i=1}^{n}\theta_0(r_i)w(r_i)$$

Lagrangian

$$\min \frac{1}{n}\sum_{i=1}^{n}\theta_1(r_i)w(r_i) + \lambda\frac{1}{n}\sum_{i=1}^{n}[\theta_1(r_i) - \theta_0(r_i)]w(r_i)$$

Differentiate wrt $w(r_i)$ gives

$$0 = \frac{1}{n}\theta_1(r_i) + \lambda\frac{1}{n}[\theta_1(r_i) - \theta_0(r_i)]$$

$$\frac{1+\lambda}{\lambda} = \frac{\theta_1(r_i)}{\theta_0(r_i)} = l(r_i)$$

If there is slack in constraint, then this holds. If $w(r_i) = 0$ then there is no slack. This says that if $l(r_i)$ is strictly increasing, at most one r_i can have slack, b/c this condition cannot possibly hold for more than one.

So question is which r_i we want to pay for. The following is the expected payment.

$$\frac{c\theta_1(r_i)}{\theta_1(r_i) - \theta_0(r_i)} = \frac{cl(r_i)}{l(r_i) - 1}$$

Note that $\theta_1(r_i) - \theta_0(r_i) > 0$ for this to satisfy limited liability. That means $l(r_i) > 1$. And the RHS is decreasing in $l(r_i)$ when that's the case. Therefore, we maximize this by setting $i = m$.

In the case with upper bound, once again we have at most one r_t with slack. Here is the expected payment with $k_1, \dots, k_n$.

$$\theta_1(r_{k_t})w(r_t) + \sum_{i=t+1}^{n} \theta_1(r_{k_i})\bar{w} \geq c + \theta_0(r_{k_t})w(r_t) + \sum_{i=t+1}^{n} \theta_0(r_{k_i})\bar{w}$$

$$[\theta_1(r_{k_t}) - \theta_0(r_{k_t})]w(r_t) \geq c + \sum_{i=t+1}^{n} [\theta_0(r_{k_i}) - \theta_1(r_{k_i})]\bar{w}$$

$$w(r_t) \geq \frac{c + \sum_{i=t+1}^{n}[\theta_0(r_{k_i}) - \theta_1(r_{k_i})]\bar{w}}{\theta_1(r_{k_t}) - \theta_0(r_{k_t})}$$

so once again you want to set t as large as possible.

But also fixing all wages except for $i, j...$ where $i > j$ but $w(r_j) = \bar{w}$ and $w(r_i) = 0$. Equate their contribution to the agent's incentive, i.e.

$$\bar{w}[\theta_1(r_i) - \theta_0(r_i)] = \gamma\bar{w}[\theta_1(r_j) - \theta_0(r_j)]$$

where $\gamma > 1$ is interpreted as we're setting $w(r_i) = \bar{w}/\gamma$ and otherwise we interpret as $w(r_j) = \bar{w}\gamma$. Either way don't violate wage constraint.

Now consider the costs. We have $\theta_1(r_i)\bar{w}$ vs. $\theta_1(r_j)\gamma\bar{w}$. What is γ?

$$\frac{[\theta_1(r_i) - \theta_0(r_i)]}{[\theta_1(r_j) - \theta_0(r_j)]} = \gamma$$

So is it true that

$$\theta_1(r_i) < \theta_1(r_j)\frac{[\theta_1(r_i) - \theta_0(r_i)]}{[\theta_1(r_j) - \theta_0(r_j)]}$$

$$\frac{\theta_1(r_i)}{\theta_1(r_i) - \theta_0(r_i)} < \frac{\theta_1(r_j)}{\theta_1(r_j) - \theta_0(r_j)}$$

$$\frac{l(r_i)}{l(r_i) - 1} < \frac{l(r_j)}{l(r_j) - 1}$$

This follows from $r_i > r_j$ and the fact that the function is decreasing. So, by switching r_i, r_j wages, we maintain the agent's incentive but lower costs to the principal. Therefore, w is increasing.

A.11 Proof of Proposition 7

proof should probably not follow Rademacher complexity, but other proofs for finite hypothesis classes – concentration inequalities for quantiles, and then union bound

A.12 Proof of Proposition 8