Who Leads and Who Follows in Strategic Classification?

Tijana Zrnic* Eric Mazumdar* S. Shankar Sastry Michael I. Jordan

University of California, Berkeley

Abstract

As predictive models are deployed into the real world, they must increasingly contend with strategic behavior. A growing body of work on strategic classification treats this problem as a *Stackelberg game*: the decision-maker "leads" in the game by deploying a model, and the strategic agents "follow" by playing their best response to the deployed model. Importantly, in this framing, the burden of learning is placed solely on the decision-maker, while the agents' best responses are implicitly treated as instantaneous. In this work, we argue that the order of play in strategic classification is fundamentally determined by the relative frequencies at which the decision-maker and the agents adapt to each other's actions. In particular, by generalizing the standard model to allow *both* players to learn over time, we show that a decision-maker that makes updates faster than the agents can reverse the order of play, meaning that the agents lead and the decision-maker follows. We observe in standard learning settings that such a role reversal can be desirable for both the decision-maker and the strategic agents. Finally, we show that a decision-maker with the freedom to choose their update frequency can induce learning dynamics that converge to Stackelberg equilibria with either order of play.

1 Introduction

Individuals interacting with a decision-making algorithm often adapt strategically to the decision rule in order to achieve a desirable outcome. While such strategic adaptation might increase the individuals' utility, it also breaks the statistical patterns that justify the decision rule's deployment. This widespread phenomenon, often known as Goodhart's law, can be summarized as: "When a measure becomes a target, it ceases to be a good measure" [39].

A growing body of work known as *strategic classification* [13, 9, 21] models this phenomenon as a two-player game in which a decision-maker "leads" and strategic agents subsequently "follow." Specifically, the decision-maker first deploys a decision rule, and the agents then take a strategic action so as to optimize their outcome according to the deployed rule, subject to natural manipulation costs. For example, a bank might make lending decisions using applicants' credit scores. Knowing this mechanism, loan applicants might sign up for a large number of credit cards in an effort to strategically increase their credit score at little effort.

One of the main goals in the literature is to develop strategy-robust decision rules; that is, rules that remain meaningful even after the agents have adapted to them. Recent work has studied strategies for finding such rules through *repeated interactions* between the decision-maker and the agents [14, 36, 2]. In particular, the decision-maker sequentially deploys different rules, and for each they observe the population's response. Under certain regularity conditions, over time the decision-maker can find the optimal solution, defined as the rule that minimizes the decision-maker's loss *after* the agents have responded to the rule.

With the emergence of online platforms such as social media and e-commerce sites, repeated interactions between decision-makers and the population have become ever more prevalent. Online platforms

*Equal contribution.

continuously monitor user behavior and update pricing algorithms, recommendation systems, and popularity rankings accordingly. Users, on the other hand, take actions to ensure favorable outcomes in the face of these updates.

A distinctive feature of online platforms is the decision-maker's dominant computational power and abundant data resources, allowing the platform to react to any change in the agents' behavior virtually instantaneously. For example, if fake news content changes over time, automated algorithms can quickly detect this and retrain the classifier to incorporate the shift. It has been observed [see, e.g., 35, 12, 10] that, when faced with such "reactive" algorithms, strategic agents tend to take actions that *anticipate* the algorithm's response. That is, through repeated interactions, agents aim to find actions that maximize the agents' utility *after* the decision-maker has responded to these actions. This suggests that the order of play in strategic interactions can in fact be *reversed*, such that the agents "lead" while the decision-maker "follows."

To give an example of such a reversed strategic interaction, consider ride-sharing platforms that deploy algorithms for determining travel fare as a function of trip length and relevant traffic conditions. These pricing mechanisms are frequently updated based on the current supply and demand, and in particular a dip in the supply of drivers triggers a surge pricing algorithm. Möhlmann and Zalmanson [35] observed that drivers occasionally coordinate a massive deactivation of drivers from the system, artificially lowering driver supply, only to get back on the platform after some time has passed and the prices have surged. In this interaction, *the drivers essentially make the first move*, while the platform's pricing algorithm reacts to their action. Other examples of users aiming to exert control over algorithms can be found in the context of social network analyses [12, 10].

In this work, we argue that the order of play in strategic classification is fundamentally tied to the relative *update frequencies* at which the decision-maker and the strategic agents adapt to each other's actions. In particular, we show that, by tuning their update frequency appropriately, the decision-maker can select the order of play in the underlying game. Furthermore, in natural settings we show that allowing the strategic agents to play first in the game can actually be preferable for *both* the decision-maker and the agents. This is contrary to the order of play previously studied in the literature, whereby the decision-maker is always assumed to make the first move.

1.1 Our contribution

To give an overview of our results, we recall some relevant game-theoretic concepts. In the existing literature strategic classification is modeled as a *Stackelberg game*. A Stackelberg game is a two-person game where one player, called the *leader*, moves first, and the other player, called the *follower*, moves second, with the possibility of adapting to the move of the leader. Previous work assumes that the decision-maker acts as the leader and the agents act as the follower. This means that the decision-maker first deploys a model, and the agents then modify their features at some cost in order to obtain a favorable outcome according to the model. The decision-maker's goal is to find the *Stackelberg equilibrium*—the model that minimizes the decision-maker's loss after the agents have optimally adapted to the model. This optimal reaction by the agents is called the *best response* to the model.

An important parameter that has been largely overlooked in existing work is the rate at which the agents re-evaluate and potentially modify their features. Most works studying the interaction between a decision-maker and strategic agents implicitly assume that, as soon as the model is updated, the data collected from strategic agents follows the best response to the currently deployed model. In the current work we do *not* assume that the agents can react instantaneously to model updates. Instead, we assume that there is a natural timescale according to which the agents adapt their features to models.

Allowing agents to adapt gradually to deployed models gives the decision-maker a new dimension upon which to act strategically. Faced with agents that adapt gradually, the decision-maker can *choose* the timescale at which they update the deployed model. In particular, they can choose a rate of updates that is *faster* than the agents' rate, or they can choose a rate that is *slower* than the agents' rate. We call decision-makers that follow a faster clock than the agents *reactive*, and if they follow a slower clock we call them *proactive*. Given that existing work on strategic classification relies on instantaneous agent responses, the previously studied decision-makers are all implicitly proactive.

Our first main result states that the decision-maker's choice of whether to be proactive or reactive fundamentally determines the order of play in strategic classification. Perhaps counterintuitively, by choosing to be *reactive* it is possible for the decision-maker to let the *agents* become the leader in the underlying Stackelberg game. Since changing the order of play changes the game's natural equilibrium concept, this choice can have a potentially important impact on the solution that the decision-maker and agents find. Throughout, we refer to the Stackelberg equilibrium when the decision-maker leads as the *decision-maker's equilibrium* and the Stackelberg equilibrium when the agents lead as the *strategic agents' equilibrium*.

Theorem 1.1 (Informal)**.** *If the decision-maker is proactive, the natural dynamics of strategic classification converge to the decision-maker's equilibrium. If the decision-maker is reactive, the natural dynamics of strategic classification converge to the strategic agents' equilibrium.*

To provide some intuition for Theorem 1.1, imagine that one player makes updates with far greater frequency than the other player. This allows the faster player to essentially converge to their best response between any two updates of the slower player. The slower player is then faced with a Stackelberg problem: they have to choose an action, expecting that the faster player will react optimally after their update. As a result, the optimal choice for the slower player is to drive the dynamics toward the Stackelberg equilibrium where they act as the leader.

It is well known (see, e.g., Section 4.5 in [1]) that under general losses, either player can prefer to lead or follow in a Stackelberg game. Our second main result shows that in classic learning problems it can be preferable for *both* the decision-maker and the agents if the agents lead in the game and the decision-maker follows.

Theorem 1.2 (Informal)**.** *Suppose that the decision-maker implements a linear or logistic regression model and the strategic agents aim to maximize their predicted score. Then, both the decision-maker and the strategic agents prefer the strategic agents' equilibrium to the decision-maker's equilibrium.*

Theorem 1.2 suggests that there are other meaningful equilibria than those previously studied in the literature. Moreover, Theorem 1.1 proves that such equilibria can naturally be achieved if the decision-maker is *reactive*. Seeing that the decision-maker's equilibrium has also been shown to imply a cost to social welfare [23, 34], our results pave the way for studying new, potentially more desirable solutions in strategic settings.

1.2 Related work

Our work builds on the growing literature on strategic classification [see, e.g., 13, 9, 21, 14, 23, 34, 27, 19, 32, and the references therein]. In these works, a decision-maker seeks to deploy a predictive model in an environment where strategic agents attempt to respond in a *post hoc* manner to maximize their utility given the model. Given this framework, a number of recent works have studied natural learning dynamics for learning models that are robust to strategic manipulation of the data [14, 23, 11, 38, 2]. Such problems have also been studied in the more general setting of performative prediction [36, 31, 8, 33, 24]. Notably, all of these works model the interaction between the decision-maker and the agents as a repeated Stackelberg game in which the decision-maker leads and the agents follow, and these roles are immutable.

Learning in Stackelberg games is itself a growing area in game-theoretic machine learning. Recent work has analyzed the asymptotic convergence of gradient-based learning algorithms to local notions of Stackelberg equilibria [16, 15, 14] assuming a fixed order of play. The emphasis has largely been on zero-sum Stackelberg games, due to their structure and relevance for min-max optimization and adversarial learning [15, 25]. Such results often build upon work on two-timescale stochastic approximations [6, 28] in which tools from dynamical systems theory are used to analyze the limiting behavior of coupled stochastically perturbed dynamical systems evolving on different timescales.

In this paper we depart from this prior work in both our problem formulation and our analysis of learning algorithms. To begin, the problem we consider is asymmetric: one player, namely the strategic agents, makes updates at a fixed frequency, while the opposing player, the decision-maker, can strategically choose their update frequency as a function of the equilibrium to which they wish to converge. Thus,

unlike prior work, the choice of timescale becomes a strategic choice on the part of the decision-maker and consequently the order of play in the Stackelberg game is not predetermined.

Our analysis of learning algorithms is also more involved than previous works since *both* the leader and follower make use of learning algorithms. Indeed, throughout our paper we assume that the population of agents is using *no-regret learning* algorithms, common in economics and decision theory [20, 29, 18, 22, 4, 5, 37]. This captures the reality that agents *gradually* adapt to the decision-maker's actions on some natural timescale. In contrast, existing literature on strategic classification and learning in Stackelberg games assume that the strategic agents or followers are always playing their best response to the leader's action.

Given this assumption on the agents' strategies, we then show how decision-makers who reason strategically about their relative update frequency can use simple learning algorithms and still be guaranteed to converge to *game-theoretically meaningful* equilibria. Our analysis complements a line of recent work on understanding gradient-based learning in continuous games, but crucially does not assume that the two players play simultaneously [see, e.g., 7, 30]). Instead, we analyze cases where *both* the agents and decision-maker learn over time, and play asynchronously.

Some of our analyses touch on ideas from online convex optimization [40], specifically derivative-free optimization [17]. Several works [14, 33] within strategic classification and performative prediction apply similar zeroth-order tools to find the decision-maker's equilibrium, but once again assuming immediate best responses to deployed models. We show that such algorithms are versatile enough to be used without such strong assumptions while still having strong convergence guarantees.

1.3 Organization

This paper is organized as follows. In Section 2 we introduce our model for studying the interaction between a decision-maker and strategic agents that adapt gradually to deployed decision rules, and formalize the concept of *reactive* and *proactive* decision-makers. In Section 3 we show that under natural assumptions, a proactive or reactive decision-maker can efficiently drive the game towards the decision-maker's or agents' equilibrium, respectively, by using simple learning rules. We follow this in Section 4 by showing how in simple learning problems inducing a certain order of play can benefit *both* the decision-maker and the agents. In Section 5 we present empirical results that corroborate our theory and emphasize how having the agents lead is more desired even in previously studied models of strategic classification. We conclude in Section 6 with a brief discussion of the questions our proposed model raises and some directions for future work.

2 Model

We start with an overview of the basic concepts and notation, and then discuss the main conceptual novelty of our work—implications of the decision-maker's and strategic agents' update frequencies.

2.1 Basic concepts and notation

Throughout we denote by $z = (x, y)$ the (feature, label) pairs corresponding to the strategic agents' data. We assume that the decision-maker chooses a model parameterized by $\theta \in \Theta \subseteq \mathbb{R}^d$, where Θ is convex and closed, and that their loss is measured via a convex loss function $\ell(z; \theta)$. The strategic agents measure loss via a function $r(z; \theta)$ and, collectively, they form a distribution in the family $\{\mathcal{P}(\mu) : \mu \in \mathcal{M} \subseteq \mathbb{R}^m\}$, where $\mathcal{M}$ is convex and closed. Here, μ denotes the agents' action.

We denote $L(\mu, \theta) = \mathbb{E}_{z\sim\mathcal{P}(\mu)}\ell(z; \theta)$, and $R(\mu, \theta) = \mathbb{E}_{z\sim\mathcal{P}(\mu)}r(z; \theta)$. With this, the agents' best response is given by $\mu_{\mathrm{BR}}(\theta) = \arg\min_\mu R(\mu, \theta)$ and the decision-maker's best response is given by $\theta_{\mathrm{BR}}(\mu) = \arg\min_\theta L(\mu, \theta)$. We assume that the best responses for both players are always unique.

If the decision-maker acts as the leader in the game, their incurred *Stackelberg risk* is equal to $\mathrm{SR}_L(\theta) = L(\mu_{\mathrm{BR}}(\theta), \theta)$. Similarly, we let $\mathrm{SR}_R(\mu) = R(\mu, \theta_{\mathrm{BR}}(\mu))$ denote the Stackelberg risk of the agents when they lead in the game. We let θ_{SE} and μ_{SE} denote the decision-maker's and strategic agents' equilibrium,

respectively: $\theta_{\mathrm{SE}} = \arg\min_\theta \mathrm{SR}_L(\theta)$ and $\mu_{\mathrm{SE}} = \arg\min_\mu \mathrm{SR}_R(\mu)$. We assume that each equilibrium is unique.

As discussed earlier, we assume that there is an underlying timescale according to which the agents re-evaluate their features. Specifically, after each time interval of fixed length, the agents observe the currently deployed model, as well as their loss according to that model, and possibly modify their features accordingly. The decision-maker, aware of the agents' timescale, can choose to be *proactive*, meaning they choose an update frequency slower than that of the agents, or *reactive*, meaning they choose a higher update frequency. This power asymmetry that allows the decision-maker to choose a timescale is characteristic of online platforms with abundant resources.

We use the term *epoch* to refer to a period between two updates of the *slower* player (which player is the slower one is up to the decision-maker). In particular, the t-th epoch starts with a single update of the slower player, followed by $\tau \in \mathbb{N}$ updates of the faster player. The rate τ is fixed.

We use θ_t and μ_t to denote the iterate of the decision-maker and the strategic agents, respectively, *at the end* of epoch t. Furthermore, for the faster player, we use double-indexing to denote the within-epoch iterates. For example, if the decision-maker is the faster player, we use $\{\theta_{t,j}\}_{j=1}^{\tau}$ to denote their iterates within epoch t. Note that $\theta_{t,\tau} \equiv \theta_t$. We also let $\bar{\theta}_t = \frac{1}{\tau}\sum_{j=1}^{\tau} \theta_{t,j}$. We adopt similar notation when the agents have a higher update frequency.

2.2 Rational agents in the face of varying update frequencies

Adopting the distinction between reactive and proactive decision-makers, it is crucial to re-evaluate what it means for the strategic agents to behave rationally. We argue that rational behavior must depend on the relative update frequencies of the decision-maker and the agents.

As a running toy example, consider a decision-maker building a model with the goal of distinguishing between spam and legitimate emails. The population of strategic agents aims to craft emails that bypass the decision-maker's spam filter. Here, μ could determine the number of words in an email, types of words used, etc. The loss $R(\mu, \theta)$ could be some decreasing function of the number of daily clicks on email content, given spam filter θ and emails crafted according to μ. In the following discussion assume that the timescales of the decision-maker and the agents have a significant separation: the decision-maker is either "significantly faster" or "significantly slower." As we will make more formal later on, our results will generally assume a sufficiently large separation between the timescales.

Proactive decision-maker. First, assume that the decision-maker is proactive, and suppose they deploy model θ. By definition, this model remains in place for a relatively long time, as observed by the agents. Then, by choosing features μ, the agents experience loss $R(\mu, \theta)$ during that period, and as a result the most rational decision is to choose features $\mu_{\mathrm{BR}}(\theta)$. In the running example, if θ is a spam filter that is in place for many months, it is rational for spammers to craft emails that are most likely to bypass filter θ. This is just the usual best response—as we alluded to earlier, when the decision-maker is proactive, our setup is similar to that of previous work.

Reactive decision-maker. Now assume that the decision-maker is reactive, and suppose the agents observe θ as the current model. Then, by setting μ, the agents do *not* experience loss $R(\mu, \theta)$. Rather, their loss is $R(\mu, \theta_{\mathrm{R}}(\mu))$, where $\theta_{\mathrm{R}}(\mu)$ denotes the decision-maker's *reaction* to the agents' choice μ. In the spam example, suppose that the decision-maker can aggregate and process data quickly, and retrains the spam filter every couple of hours. Moreover, suppose that the spammers adapt their emails only once per week. Then, the agents' loss after choosing μ (evaluated weekly) is determined by the number of clicks allowed by the updated filter $\theta_{\mathrm{R}}(\mu)$, *not* the old filter θ. Therefore, if the agents could predict $\theta_{\mathrm{R}}(\mu)$, the agents' optimal decision would be to choose $\arg\min_\mu R(\mu, \theta_{\mathrm{R}}(\mu))$. In other words, rather than choose the best response to θ, rational agents interacting with a reactive decision-maker would choose μ so that it triggers the *best possible reaction* from θ.

We formalize this intuitive behavior by assuming that the agents are *no-regret* learners [37]. This essentially means that their average regret vanishes as the number of actions grows. More formally, we assume the following behavior depending on the relative update frequencies:

- If the decision-maker is proactive, then for any θ_t, the agents' strategy ensures:

$$\frac{1}{\tau}\sum_{j=1}^{\tau}\mathbb{E}R(\mu_{t,j},\theta_t) - \min_{\mu} R(\mu,\theta_t) \to 0 \text{ as } \tau \to \infty. \tag{A1}$$

- If the decision-maker is reactive, then for any response function $\theta_{\mathrm{R}}(\mu)$, the agents' strategy ensures:

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}R(\mu_t,\theta_{\mathrm{R}}(\mu_t)) - \min_{\mu} R(\mu,\theta_{\mathrm{R}}(\mu)) \to 0 \text{ as } T \to \infty, \tag{A2}$$

whenever such a strategy exists. If the agents' loss is convex, the first condition can be satisfied by simple gradient descent. In fact, gradient descent would typically imply an even stronger guarantee, namely the convergence of the iterates, $\mu_{t,\tau} \to \mu_{\mathrm{BR}}(\theta_t)$. The second condition can be satisfied by various bandit strategies if $R(\mu,\theta_{\mathrm{R}}(\mu))$ is Lipschitz and $\mathcal{M}$ is bounded (and we will impose these conditions explicitly in the following section). That said, it seems hardly suitable to assume that the agents run a well-specified optimization procedure. For this reason, we will for the most part avoid making explicit algorithmic assumptions on the agents' strategy and our main takeaways will only rely on rational agent behavior in the limit, as in equations (A1) and (A2).

3 Learning dynamics

In this section, we study the limiting behavior of the interaction between the decision-maker and the strategic agents. We show that, by running classical optimization algorithms, the decision-maker can drive the interaction to a Stackelberg equilibrium with either player acting as the leader.

3.1 Convergence to decision-maker's equilibrium

In general, we do not expect the decision-maker to be able to compute derivatives of the function SR_L. For this reason, to achieve convergence to the decision-maker's equilibrium, we consider running a derivative-free method. One such solution is the "gradient descent without a gradient" algorithm of Flaxman et al. [17]. Past work [14, 33] also considers this algorithm with the goal of optimizing SR_L, but it assumes instantaneous agent responses. In other words, it assumes query access to SR_L directly, while we consider perturbations due to imperfect agent responses. It is worth noting that, under further assumptions, one could apply more efficient two-stage approaches [33, 24] that approximate the gradients of SR_L by first estimating the best-response map μ_{BR}.

Specifically, we let the decision-maker run the following update:

$$\phi_{t+1} = \Pi_{\Theta}(\phi_t - \eta_t \frac{d}{\delta} L(\bar{\mu}_t, \phi_t + \delta u_t)u_t), \text{ where } u_t \sim \mathrm{Unif}\left(\mathcal{S}^{d-1}\right). \tag{1}$$

Here, Π_{Θ} denotes the Euclidean projection, $\mathrm{Unif}\left(\mathcal{S}^{d-1}\right)$ denotes the uniform distribution on the unit sphere in $\mathbb{R}^d$, η_t is a non-increasing step size sequence, and $\delta > 0$ is a fixed hyperparameter. The deployed model in the t-th epoch is set as $\theta_t = \phi_t + \delta u_t$.[1]

We provide convergence guarantees assuming that the decision-maker's Stackelberg risk SR_L is convex. While this condition doesn't follow from convexity of the loss $\ell(z;\theta)$ alone, previous work has established conditions for convexity of this objective for different learning problems and agent utilities [14, 33]. For example, in the linear and logistic regression examples discussed in the following section, the decision-maker's Stackelberg risk will be convex.

[1] Technically, this assumes that we can deploy a model in a δ-ball around Θ. Another solution would be to use a projection onto a small contraction of Θ in equation (1). This is a minor technical hurdle common in the literature. The rate in Theorem 3.1 is unaffected by the choice of solution to this technical point.

Theorem 3.1. *Denote diam$(\Theta) = D_\Theta$, and suppose that $|L(\mu,\theta)| \le B$ for all μ,θ. Furthermore, suppose that SR_L is convex and β-Lipschitz and $L(\mu,\theta)$ is β_μ-Lipschitz in the first entry for all θ. Then, if the decision-maker runs update* (1) *with $\eta_t = \eta_0 d^{-\frac{1}{2}} t^{-\frac{3}{4}}$ and $\delta = \delta_0 d^{\frac{1}{2}} T^{-1/4}$, it holds that*

$$\sum_{t=1}^{T} (\mathbb{E}[\mathrm{SR}_L(\theta_t)] - \mathrm{SR}_L(\theta_{\mathrm{SE}})) \le \left(\frac{D_\Theta^2}{2\eta_0} + \frac{2B^2}{\delta_0^2}\right)\sqrt{d}T^{3/4} + \beta_\mu D_\Theta \sum_{t=1}^{T} \mathbb{E}\|\bar{\mu}_t - \mu_{\mathrm{BR}}(\theta_t)\|_2.$$

Moreover, assuming that the agents are rational (A1) *and $\mathcal{M}$ is compact, we have*

$$\lim_{\tau\to\infty} \sum_{t=1}^{T} (\mathbb{E}[\mathrm{SR}_L(\theta_t)] - \mathrm{SR}_L(\theta_{\mathrm{SE}})) \le \left(\frac{D_\Theta^2}{2\eta_0} + \frac{2B^2}{\delta_0^2}\right)\sqrt{d}T^{3/4}. \tag{2}$$

Remark 3.2. *For Theorem 3.1, we assume that the agents are rational in a relatively weak sense, by assuming no-regret behavior. Often, however, we expect the agents' strategy to achieve* iterate convergence*, and not just vanishing regret. More precisely, it makes sense to expect $\mu_{t,\tau} \to \mu_{\mathrm{BR}}(\theta_t)$ as $\tau \to \infty$. For example, this guarantee is achieved by gradient descent in a variety of settings. In that case, the decision-maker can simply use the* last *iterate instead of the average one:*

$$\phi_{t+1} = \Pi_\Theta(\phi_t - \eta_t \frac{d}{\delta} L(\mu_t, \phi_t + \delta u_t) u_t), \text{ where } u_t \sim \mathrm{Unif}\left(\mathcal{S}^{d-1}\right). \tag{3}$$

Similarly, $\mathbb{E}\|\bar{\mu}_t - \mu_{\mathrm{BR}}(\theta_t)\|_2$ would be replaced by $\mathbb{E}\|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\|_2$ in the bound of Theorem 3.1.

In some cases, the additional regret due to imperfect agent responses does not alter the asymptotic rate at which the decision-maker accumulates regret even if the epoch length τ is constant and does not grow with T. To illustrate this point, we consider strategic agents that follow the gradient-descent direction on a possibly nonconvex objective with enough curvature. More precisely, we assume that for all θ, $R(\mu,\theta)$ satisfies the Polyak-Łojasiewicz (PL) condition:

$$\gamma(R(\mu,\theta) - \min_{\mu\in\mathcal{M}} R(\mu,\theta)) \le \frac{1}{2}\|\nabla_\mu R(\mu,\theta)\|_2^2,$$

for some parameter $\gamma > 0$. Suppose that the agents' update is computed as:

$$\mu_{t,j+1} = \mu_{t,j} - \eta_\mu \nabla_\mu R(\mu_{t,j}, \theta_t), \tag{4}$$

where $\eta_\mu > 0$ is a constant step size and $\mu_{t,0} = \mu_{t-1,\tau}$. In this case, gradient descent achieves last-iterate convergence and hence we assume that the decision-maker uses the update in equation (3).

Theorem 3.3. *Assume the conditions of Theorem 3.1. In addition, suppose that $R(\mu,\theta)$ is β_μ^R-smooth in μ for all θ and satisfies the PL condition with parameter γ, and $\mu_{\mathrm{BR}}(\theta)$ is β_{BR}-Lipschitz in θ. Assume that the strategic agents run update* (4) *with $\eta_\mu < \frac{1}{\beta_\mu^R}$. Further, suppose the epoch length is chosen so that $\tau > \log(\beta_\mu^R/\gamma)/\log(1/(1-\gamma\eta_\mu))$. Then, for some constant $\alpha(\tau) \in (0,1)$, we have*

$$\sum_{t=1}^{T} \mathbb{E}\|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\|_2 \le \frac{\|\mu_0 - \mu_{\mathrm{BR}}(\theta_0)\|_2 + \frac{4\beta_{\mathrm{BR}} B\eta_0}{\delta_0}\sqrt{T}}{1-\alpha(\tau)}.$$

Therefore, the decision-maker's regret is $O(\sqrt{d}T^{3/4})$ even with a constant epoch length. This result crucially depends on the fact that the optimization problems that the agents solve in neighboring epochs are coupled through $\mu_{t,0} = \mu_{t-1,\tau}$. If $\mu_{t,0}$ were reinitialized arbitrarily in each epoch, the extra regret would be linear in T given constant epoch length.

By using standard tools from the stochastic approximation literature [6], in the Appendix we additionally provide convergence to local optima for the update (3) when SR_L is possibly nonconvex.

3.2 Convergence to strategic agents' equilibrium

Now we analyze the case when the decision-maker is reactive. Given a large enough gap in update frequencies—that is, a large enough epoch length τ—the decision-maker can converge to their best response to the current iterate μ_t between any two actions of the agents. The most natural choice for achieving this is to run standard gradient descent, $\theta_{t,k+1} = \theta_{t,k} - \eta_k \nabla_\theta L(\mu_t, \theta_{t,k})$. In what follows we provide asymptotic guarantees assuming that the decision-maker runs any algorithm that achieves iterate convergence. This condition can be satisfied by gradient descent in a variety of settings. Formally, we assume that for any fixed μ_t, the decision-maker's strategy ensures

$$\|\theta_{t,\tau} - \theta_{\mathrm{BR}}(\mu_t)\|_2 \to_p 0, \tag{5}$$

as $\tau \to \infty$. Here, $\to_p$ denotes convergence in probability.

We first observe that, in the limit as τ grows, the agents' accumulated risk is equal to their accumulated *Stackelberg risk* at all the actions played so far. This simply follows by continuity.

Lemma 3.4. *Suppose that the decision-maker achieves iterate convergence* (5) *and R is continuous in the second argument. Then, for all $T \in \mathbb{N}$, $\lim_{\tau\to\infty} \sum_{t=1}^{T} \mathbb{E}R(\mu_t, \theta_t) = \sum_{t=1}^{T} \mathbb{E}\mathrm{SR}_R(\mu_t)$.*

In other words, in every epoch the agents essentially play a Stackelberg game in which they lead and the decision-maker follows. This holds regardless of whether the agents behave rationally. If they do behave rationally (condition (A2)), we show that both the agents' and the decision-maker's average regret with respect to $(\mu_{\mathrm{SE}}, \theta_{\mathrm{BR}}(\mu_{\mathrm{SE}}))$ vanishes if the agents' updates are continuous. To formalize this, suppose that for all $t \in \mathbb{N}$, the agents set $\mu_{t+1} = D_{t+1}(\mu_1, \theta_1, \ldots, \mu_t, \theta_t, \xi_{t+1})$, where ξ_{t+1} is a random variable independent of $\{(\mu_i, \theta_i)\}_{i\le t}$. We include ξ_{t+1} as an input to allow randomized strategies. Then, we will say that the agents' updates are *continuous* if D_{t+1} is continuous in the first $2t$ coordinates for all $t \in \mathbb{N}$.

Theorem 3.5. *Suppose that the agents' updates are continuous and rational* (A2)*, and that $\mathcal{M}$ is compact. Further, suppose that the decision-maker achieves iterate convergence* (5) *and SR_R and SR_L are Lipschitz. Then, it holds that*

$$\lim_{T\to\infty} \lim_{\tau\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\mathrm{SR}_R(\mu_t) - \mathrm{SR}_R(\mu_{\mathrm{SE}}) = 0, \quad \lim_{T\to\infty} \lim_{\tau\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}L(\mu_t, \theta_t) - L(\mu_{\mathrm{SE}}, \theta_{\mathrm{BR}}(\mu_{\mathrm{SE}})) = 0.$$

4 Preferred order of play

While we have shown that the decision-maker can tune their update frequency to achieve either order of play in the Stackelberg game, it remains to understand which order of play is preferable for the decision-maker and the strategic agents. In the following examples, we illustrate that in classic learning settings both players can prefer the order when the *agents lead*. This suggests that the natural and overall more desirable order of play is sometimes reversed compared to the order usually studied.

At first, it might seem counterintuitive that the decision-maker could prefer to follow. To get some intuition for why following might be preferred to leading, recall that in zero-sum games *following is never worse*. In particular, suppose $R(\mu, \theta) = -L(\mu, \theta)$. Then, the basic min-max inequality says

$$L(\mu_{\mathrm{SE}}, \theta_{\mathrm{BR}}(\mu_{\mathrm{SE}})) = \max_{\mu} \min_{\theta} L(\mu, \theta) \le \min_{\theta} \max_{\mu} L(\mu, \theta) = L(\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}), \theta_{\mathrm{SE}}),$$

with equality if and only if a *Nash* equilibrium exists. Therefore, if a Nash equilibrium does not exist, following is strictly preferred.

Since strategic classification is typically not a zero-sum game, we look at two common learning problems and analyze the preferred order of play.

4.1 Linear regression

Suppose that the agents' non-strategic data, (x_0, y), where x_0 is a feature vector and y the outcome, is generated according to

$$x_0 \sim N(0, I_d),\ y = x_0^\top \beta + \xi,\ \xi \sim N(0, \sigma^2),$$

where $\beta \in \mathbb{R}^d$ is an arbitrary fixed vector. We denote the joint distribution of (x_0, y) by $\mathcal{P}(0)$.

Recall that we use z to denote the pair (x, y). Suppose that the decision-maker runs standard linear regression with the squared loss:

$$\ell(z; \theta) = \frac{1}{2}(y - x^\top \theta)^2.$$

The agents aim to maximize their predicted outcome, $r(z; \theta) = -\theta^\top x$, subject to a fixed budget on feature manipulation—they can move to any x at distance at most B from their original features x_0: $\|x - x_0\|_2 \le B$. A similar model is considered by Chen et al. [11]. More precisely, we let $\mathcal{M} = \{\mu \in \mathbb{R}^d : \|\mu\|_2 \le B\}$ and define $\mathcal{P}(\mu)$ to be the distribution of (x, y), where $(x_0, y) \sim \mathcal{P}(0)$ and $x = x_0 + \mu$. Then, $R(\mu, \theta) = \mathbb{E}_{z\sim\mathcal{P}(\mu)} r(z; \theta) = -\mu^\top \theta$ and $L(\mu, \theta) = \mathbb{E}_{z\sim\mathcal{P}(\mu)} \ell(z; \theta)$.

We prove that both the decision-maker and the agents prefer the agents' equilibrium.

Proposition 4.1. *Assume the linear regression setup described above. Then, we have*

$$\frac{\sigma^2}{2} + \frac{\|\beta\|_2^2 \min(1, B)^2}{2(1 + \min(1, B)^2)} = L(\mu_{\mathrm{SE}}, \theta_{\mathrm{BR}}(\mu_{\mathrm{SE}})) \le \mathrm{SR}_L(\theta_{\mathrm{SE}}) = \frac{\sigma^2}{2} + \frac{\|\beta\|_2^2 B^2}{2(1 + B^2)},$$
$$-\frac{\|\beta\|_2 \min(1, B)}{1 + \min(1, B)^2} = \mathrm{SR}_R(\mu_{\mathrm{SE}}) \le R(\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}), \theta_{\mathrm{SE}}) = -\frac{\|\beta\|_2 B}{1 + B^2}.$$

When $B \le 1$, the losses implied by the two scenarios are the same, while when $B > 1$, having the agents lead is strictly better for both players. Moreover, the strategic agents' manipulation cost is no higher when they lead: $\|\mu_{\mathrm{SE}}\|_2 \le \|\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}})\|_2$.

4.2 Logistic regression

Next we consider a classification example. Suppose that the non-strategic data is generated as

$$y \sim \mathrm{Bern}\,(p) \ \text{ and } x_0|y \sim N((2y - 1)\alpha, \sigma^2 I).$$

In other words, $x_0|y = 1 \sim N(\alpha, \sigma^2 I)$ and $x_0|y = 0 \sim N(-\alpha, \sigma^2 I)$. We denote the joint distribution over (x_0, y) by $\mathcal{P}(0)$.

We assume that the decision-maker trains a logistic regression classifier:

$$\ell(z; \theta) = -yx^\top \theta + \log(1 + e^{x^\top \theta}).$$

The agents with $y = 0$ can manipulate their features to increase the probability of being positively labeled. A similar setup is considered by Dong et al. [14]. As in the previous example, the agents have a limited budget to change their features: if their non-strategic features are x_0, they can move to any x which is at distance at most B from x_0, $\|x - x_0\|_2 \le B$. Thus, we set $\mathcal{M} = \{\mu \in \mathbb{R}^d : \|\mu\|_2 \le B\}$ and denote by $\mathcal{P}(\mu)$ the joint distribution of (x, y) where $(x_0, y) \sim \mathcal{P}(0)$ and $x = x_0 + \mu \mathbf{1}\{y = 0\}$. We impose a mild condition that $B \le \|\alpha\|_2$, so that the population whose true outcome is 0 cannot move closer to α than $-\alpha$. We let $R(\mu, \theta) = -\mu^\top \theta$ and $L(\mu, \theta) = \mathbb{E}_{z\sim\mathcal{P}(\mu)} \ell(z; \theta)$.

Proposition 4.2. *Assume the logistic regression setup described above. Then, we have*

$$L(\mu_{\mathrm{SE}}, \theta_{\mathrm{BR}}(\mu_{\mathrm{SE}})) \le \mathrm{SR}_L(\theta_{\mathrm{SE}}) \text{ and } \mathrm{SR}_R(\mu_{\mathrm{SE}}) \le R(\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}), \theta_{\mathrm{SE}}).$$

There exist configurations of parameters such that the inequalities in Proposition 4.2 are strict, meaning that both players strictly prefer the agents to lead. We illustrate this empirically. In Figure 1 we plot the difference in risk between the two equilibria for the decision-maker and the agents, for varying B and p. For large p and small B, we see no difference between the equilibria. However, as p decreases and B increases, it becomes suboptimal for both players if the decision-maker leads.

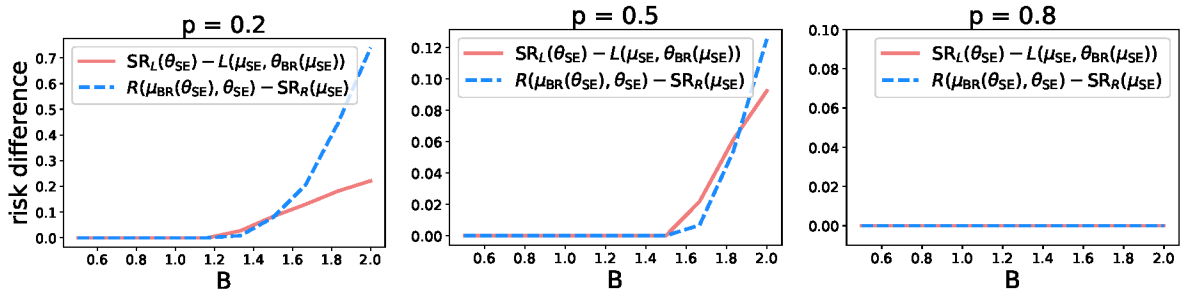

Figure 1: Difference in decision-maker's and agents' risk implied by the two Stackelberg equilibria, for different values of B and p. We set $d = 1, \alpha = 2, \sigma = 1$.

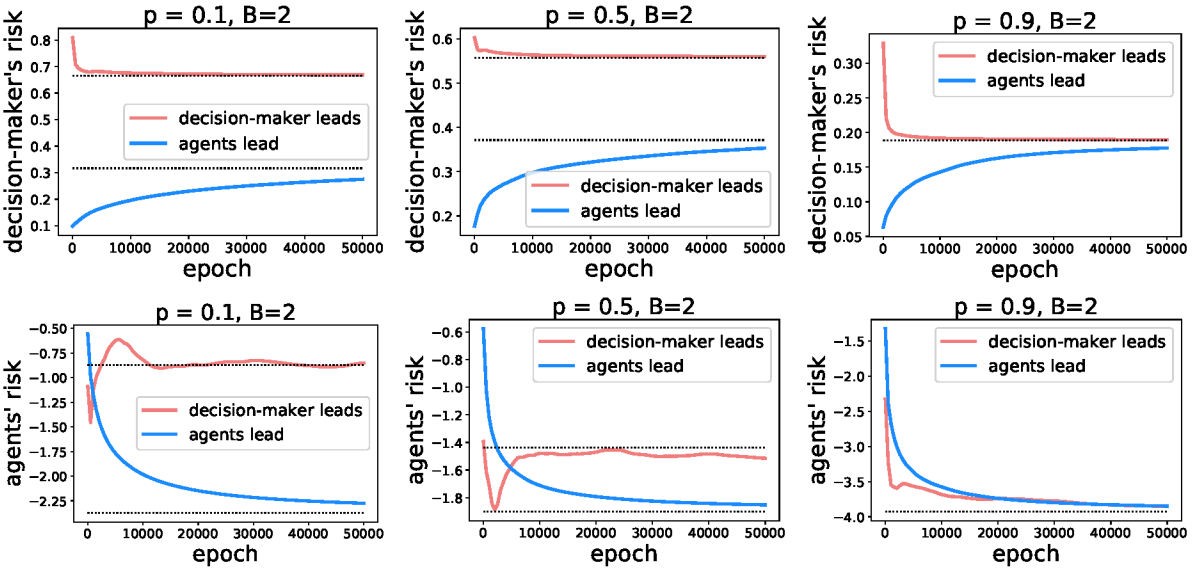

Figure 2: Decision-maker's and agents' average running risk for varying p and $B = 2$. The dotted lines denote the loss at the respective equilibria. For $p = 0.9$, the decision-maker's equilibrium and the agents' equilibrium coincide and hence both curves converge to the same value.

5 Experiments

As proof of concept, we demonstrate our theoretical findings empirically in a simulated logistic regression setting. In the first set of experiments, we adopt the model from Section 4.2 where agents are constrained in how they modify their features. In the second set of experiments we adopt a model more akin to that of Dong et al. [14] where the negatively classified agents are penalized from deviating from their true features. The details of all numerical results are deferred to the end of this section.

In both settings, first we let the decision-maker lead and the agents follow, and then we switch the roles. For both orders of play, the slower player runs the derivative-free update (3), and the faster player runs standard (projected) gradient descent. To be able to analyze the long-run behavior, we also numerically approximate the Stackelberg risks of the decision-maker and the strategic agents and find the global minima which correspond to the decision-maker's and agents' equilibria respectively.

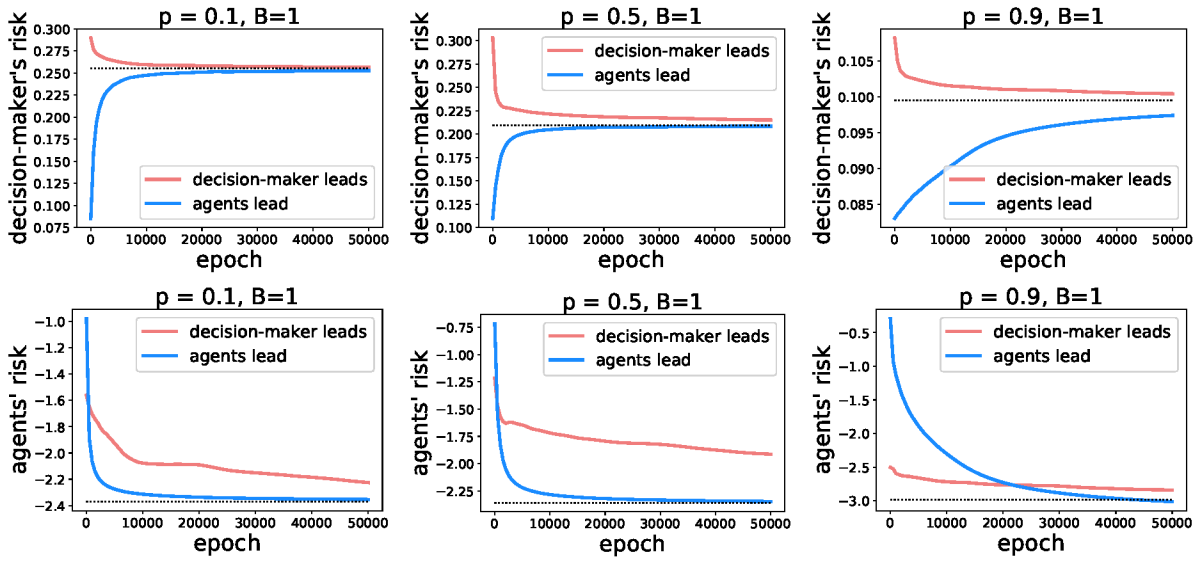

Figure 3: Decision-maker's and agents' average running risk for varying p and $B = 1$. The dotted lines denote the loss at the respective equilibria. For all values of p, the decision-maker's equilibrium and the agents' equilibrium coincide and hence both curves converge to the same value.

5.1 Agents with constraints

To begin, we verify our theoretical findings from Section 4.2. We generate 100 samples, fix $\alpha = 2$, $\sigma = 1$, $d = 1$, and vary B and p. We run the interaction for a total of $T = 50000$ epochs, with each epoch of length $\tau = 200$.

In Figure 2 and Figure 3 we plot the decision-maker's and the agents' average running risk against the number of epochs, for the two different orders of play, for $B = 2$ and $B = 1$, respectively. For $p \in \{0.1, 0.5\}$ and $B = 2$, we observe a clear gap between leading and following, the agents leading being the preferred order for both players. For $p = 0.9$ or $B = 1$, the two equilibria coincide asymptotically; however, generally we find that the two players still prefer the agents to lead even after a finite number of epochs.

5.2 Agents with costly deviations

In this section, we verify our findings on a model where the decision-maker's problem is the same logistic regression problem posed in Section 4.2, but the strategic agents are penalized for deviating from their true features. In particular, the agents' risk R takes the form:

$$R(\mu, \theta) = \frac{\lambda}{2}\|\mu\|^2 - \mu^T\theta.$$

We note that although this setup is conceptually very similar to that in Section 4.2 (increasing λ can be seen as shrinking the constraint set), it allows us to highlight that the experimental results are not caused by interactions with the constraints. Further, this setup is more readily comparable to previous models studied in, e.g., [14].

We generate 100 samples in $\mathbb{R}^2$, fix $\alpha = 1.5[1, 1]^T$ and $\sigma = 1$, and vary λ and p. We run the interaction for a total of $T = 50000$ epochs, with each epoch of length $\tau = 100$. In Figure 4 and Figure 5 we plot the decision-maker's and the agents' average running risk against the number of epochs, for the two different orders of play and for $\lambda = 1$ and $\lambda = 20$ respectively.

In Figure 4 we observe a gap between the decision-maker's risk at their Stackelberg equilibrium and at the agents', and see that the decision-maker consistently achieves a lower risk when the agents lead.

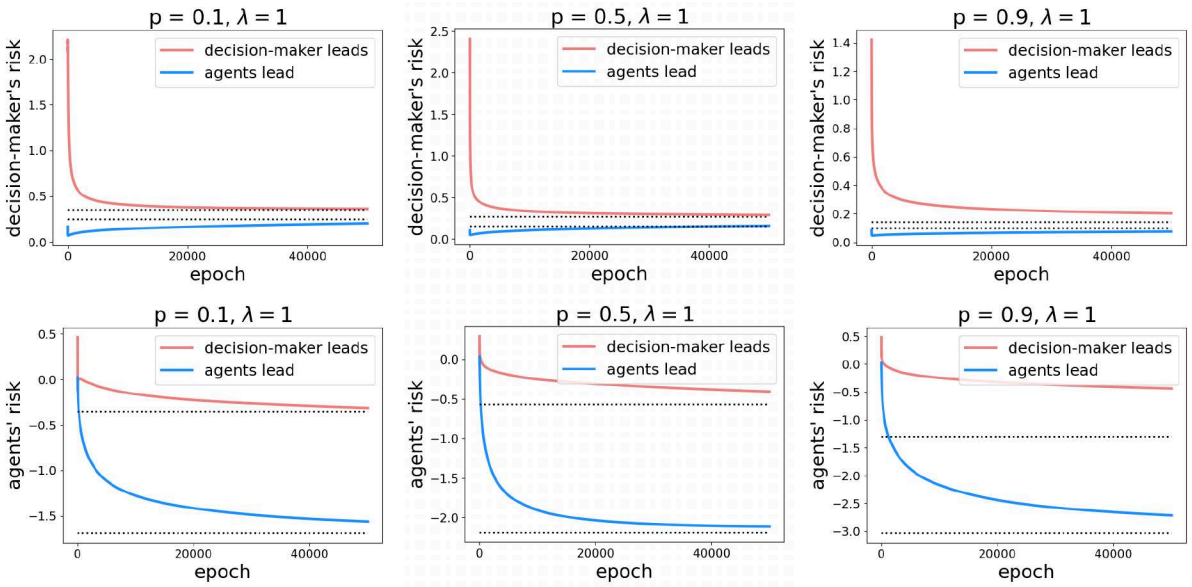

Figure 4: Decision-maker's and agents' average running risk for varying p and $\lambda = 1$. The dotted lines denote the loss at the respective equilibria.

Further, we note that agents consistently prefer leading, meaning that *both* the decision-maker and agents prefer if the order of play is flipped. In Figure 5 we observe that as λ and p increase, the gap between the two equilibria shrinks and disappears entirely when $p = 0.9$ and $\lambda = 20$. This is similar to the behavior seen in the constrained agent problem where shrinking the constraint set can give rise to Nash equilibria where neither player strictly prefers leading or following.

6 Discussion

We have shown how the consideration of update frequencies allows natural learning dynamics to converge to Stackelberg equilibria where either player can act as the leader. Moreover, we observed that the previously unexplored order of play in which the *agents lead* can result in lower risk for both players. We have only begun to understand the implications of reversing the order of play in strategic classification, and update frequencies in general, and many questions remain open for future work.

In social settings, there are many considerations and concerns beyond minimizing risk. While our preliminary observations suggest that reversing the order of play might have benefits, we have yet to fully understand the impact of this reversed order on the population interacting with the model. That said, we do not propose a new type of interaction; real-world decision-making algorithms already possess, and employ, the power to be reactive. Our new framework is simply flexible enough to capture the difference between proactive and reactive decision-makers.

Furthermore, we assume that the agents act on a fixed timescale. Sometimes it is possible for the agents to choose their timescale *strategically*. In that case, there is first a "meta-game" between the decision-maker and the agents, as they might compete for the leader/follower role. For example, if both prefer to lead, then both might aim to make slow updates to reach the leader position; perhaps surprisingly, this incentive might prevent any interaction at all.

Finally, we study order-of-play preferences only in linear/logistic regression with linear agent utilities. There are many other learning settings and classes of agents' utilities and costs in the literature, and going forward it is important to obtain general conditions when leading (or following) is preferable.

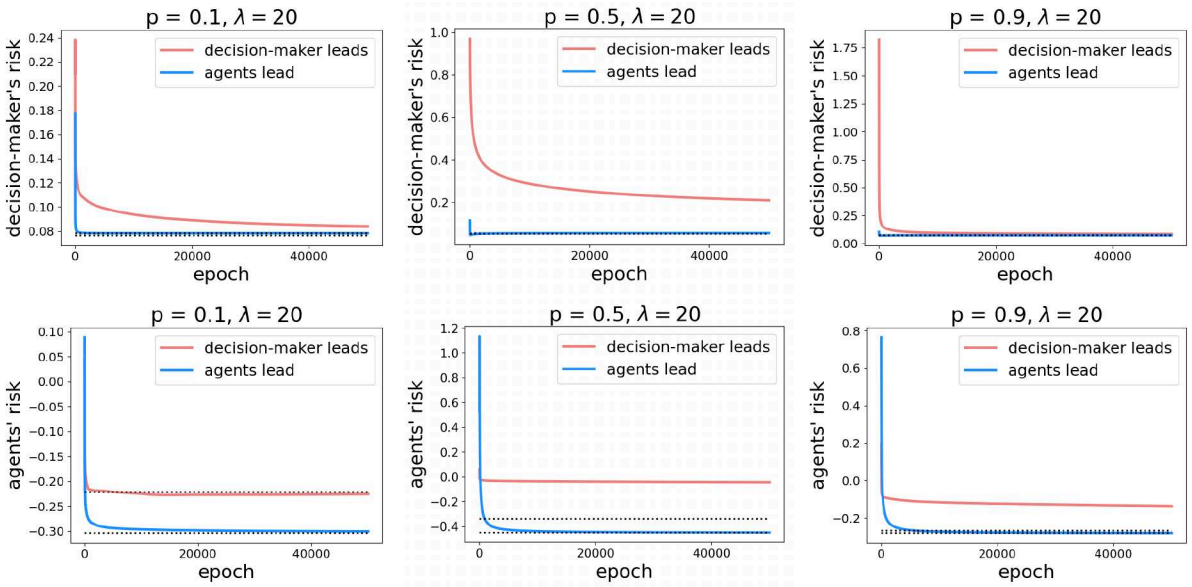

Figure 5: Decision-maker's and agents' average running risk for varying p and $\lambda = 20$. The dotted lines denote the loss at the respective equilibria. For $p = 0.9$, the decision-maker's equilibrium and the agents' equilibrium coincide and hence both curves converge to the same value.

Acknowledgements

We thank Moritz Hardt for an inspiring discussion and helpful feedback on this project. We wish to acknowledge support from the Office of Naval Research under the Vannevar Bush Fellowship program and support from HICON-LEARN (Design of HIgh CONfidence LEARNing-Enabled Systems), Defense Advanced Research Projects Agency award number FA8750-18-C-0101.

References

[1] Tamer Başar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory*. SIAM, 1998.

[2] Yahav Bechavod, Katrina Ligett, Steven Wu, and Juba Ziani. Gaming helps! Learning from strategic interactions in natural dynamics. In *International Conference on Artificial Intelligence and Statistics*, pages 1234–1242, 2021.

[3] Michel Benaïm. Dynamics of stochastic approximation algorithms. In *Seminaire de probabilites XXXIII*, pages 1–68. Springer, 1999.

[4] Avrim Blum, Eyal Even-Dar, and Katrina Ligett. Routing without regret: On convergence to Nash equilibria of regret-minimizing algorithms in routing games. In *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, pages 45–52, 2006.

[5] Avrim Blum, MohammadTaghi Hajiaghayi, Katrina Ligett, and Aaron Roth. Regret minimization and the price of total anarchy. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 373–382, 2008.

[6] Vivek .S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

[7] Mario Bravo, David Leslie, and Panayotis Mertikopoulos. Bandit learning in concave n-person games. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 5666–5676, 2018.

[8] Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. *arXiv preprint arXiv:2011.03885*, 2020.

[9] Michael Brückner, Christian Kanzow, and Tobias Scheffer. Static prediction games for adversarial learning problems. *Journal of Machine Learning Research*, 13(Sep):2617–2654, 2012.

[10] Jenna Burrell, Zoe Kahn, Anne Jonas, and Daniel Griffin. When users control the algorithms: Values expressed in practices on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, 3:19, 2019.

[11] Yiling Chen, Yang Liu, and Chara Podimata. Learning strategy-aware linear classifiers. In *Advances in Neural Information Processing Systems*, volume 33, pages 15265–15276, 2020.

[12] Kelley Cotter. Playing the visibility game: How digital influencers and algorithms negotiate influence on instagram. *New Media & Society*, 21(4):895–913, 2019.

[13] Nilesh Dalvi, Pedro Domingos, Sumit Sanghai, and Deepak Verma. Adversarial classification. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 99–108, 2004.

[14] Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70. ACM, 2018.

[15] Tanner Fiez and Lillian J Ratliff. Local convergence analysis of gradient descent ascent with finite timescale separation. In *International Conference on Learning Representations*, 2021.

[16] Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Implicit learning dynamics in Stackelberg games: Equilibria characterization, convergence analysis, and empirical study. In *International Conference on Machine Learning*, pages 3133–3144, 2020.

[17] Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Symposium on Discrete Algorithms*, pages 385–394, 2005.

[18] Dean P Foster and Rakesh V Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40, 1997.

[19] Nika Haghtalab, Nicole Immorlica, Brendan Lucier, and Jack Z. Wang. Maximizing welfare with incentive-aware evaluation mechanisms. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pages 160–166, 2020.

[20] James Hannan. Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games*, 21(39):97, 1957.

[21] Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the ACM Conference on Innovations in Theoretical Computer Science*, pages 111–122, 2016.

[22] Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, 2000.

[23] Lily Hu, Nicole Immorlica, and Jennifer Wortman Vaughan. The disparate effects of strategic manipulation. In *Proceedings of the 2nd ACM Conference on Fairness, Accountability, and Transparency*, pages 259–268, 2019.

[24] Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: performative gradient descent. *arXiv preprint arXiv:2102.07698*, 2021.

[25] Chi Jin, Praneeth Netrapalli, and Michael Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In *Proceedings of the 37th International Conference on Machine Learning*, pages 4880–4889, 2020.

[26] Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 795–811, 2016.

[27] Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? In *Proceedings of the ACM Conference on Economics and Computation (EC)*, pages 825–844, 2019.

[28] Vijay R. Konda and John N. Tsitsiklis. Convergence rate of linear two-time-scale stochastic approximation. *The Annals of Applied Probability*, 14(2):796 – 819, 2004.

[29] Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

[30] Eric Mazumdar, Lillian J. Ratliff, and S. Shankar Sastry. On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, 2020.

[31] Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. In *Advances in Neural Information Processing Systems*, volume 33, pages 4929–4939, 2020.

[32] John Miller, Smitha Milli, and Moritz Hardt. Strategic classification is causal modeling in disguise. In *International Conference on Machine Learning*, pages 6917–6926, 2020.

[33] John Miller, Juan C Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. *arXiv preprint arXiv:2102.08570*, 2021.

[34] Smitha Milli, John Miller, Anca D Dragan, and Moritz Hardt. The social cost of strategic classification. In *Proceedings of the 2nd ACM Conference on Fairness, Accountability, and Transparency*, pages 230–239, 2019.

[35] Marieke Möhlmann and Lior Zalmanson. Hands on the wheel: Navigating algorithmic management and Uber drivers'. In *Proceedings of the international conference on information systems (ICIS)*, pages 10–13, 2017.

[36] Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609, 2020.

[37] Tim Roughgarden. Algorithmic game theory. *Communications of the ACM*, 53(7):78–86, 2010.

[38] Yonadav Shavit, Benjamin Edelman, and Brian Axelrod. Causal strategic linear regression. In *International Conference on Machine Learning*, pages 8676–8686, 2020.

[39] Marilyn Strathern. Improving ratings: Audit in the British university system. *European Review*, 5 (3):305–321, 1997.

[40] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pages 928–936, 2003.

A Proofs

Lemma A.1. *Suppose that $\mathcal{M}$ is compact. If the decision-maker is proactive and the strategic agents' actions satisfy condition* (A1)*, then*

$$\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{j=1}^{\tau}\mathbb{E}\|\mu_{j,\tau}-\mu_{\mathrm{BR}}(\theta_t)\|_2=0.$$

Similarly, if the decision-maker is reactive and

$$\lim_{T\to\infty}\lim_{\tau\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\mathrm{SR}_R(\mu_t)-\mathrm{SR}_R(\mu_{\mathrm{SE}})=0,$$

then

$$\lim_{T\to\infty}\lim_{\tau\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\|\mu_t-\mu_{\mathrm{SE}}\|_2=0.$$

Proof. We will prove the second statement; the proof of the first statement is completely analogous.

By the uniqueness of μ_{SE} and compactness of $\mathcal{M}$, notice that for all μ and $\epsilon>0$ such that $\|\mu-\mu_{\mathrm{SE}}\|_2\geq\epsilon$, we have $\mathrm{SR}_R(\mu)-\mathrm{SR}_R(\mu_{\mathrm{SE}})\geq\delta(\epsilon)>0$, for some $\delta(\epsilon)$. We will use this observation to argue that, if $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\|\mu_t-\mu_{\mathrm{SE}}\|_2\not\to 0$, then that must imply positive regret in the limit, which concludes the proof by contradiction.

Denote $\mathrm{dist}_t=\lim_{\tau\to\infty}\mathbb{E}\|\mu_t-\mu_{\mathrm{SE}}\|_2$, and suppose that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathrm{dist}_t\neq 0.$$

Then, that implies that for every $\epsilon>0$, there is a sequence $\{a_k\}_{k=1}^{\infty}$ such that $\frac{1}{a_k}\sum_{t=1}^{a_k}\mathrm{dist}_t>\epsilon$ for all k. Fix $0<\epsilon'<\epsilon$, and denote $p_k=\frac{1}{a_k}|\{t\leq a_k:\mathrm{dist}_t>\epsilon'\}|$. Then, we have

$$\epsilon<\frac{1}{a_k}\sum_{t=1}^{a_k}\mathrm{dist}_t\leq p_k D_{\mathcal{M}}+\epsilon',$$

where $D_{\mathcal{M}}=\max_{\mu,\mu'\in\mathcal{M}}\|\mu-\mu'\|_2$. Therefore, $p_k\geq\frac{\epsilon-\epsilon'}{D_{\mathcal{M}}}>0$. This shows that in the sum $\frac{1}{a_k}\sum_{t=1}^{a_k}\mathrm{dist}_t$ there is a *constant* fraction of terms outside a ball of radius ϵ' around μ_{SE}, in expectation. Fix one such term dist_{t^*}. Then, we know

$$\epsilon'\leq\mathrm{dist}_{t^*}\leq\lim_{\tau\to\infty}\mathbb{P}\{\|\mu_{t^*}-\mu_{\mathrm{SE}}\|_2\geq\epsilon'/2\}D_{\mathcal{M}}+\epsilon'/2.$$

Therefore, we can conclude that $\lim_{\tau\to\infty}\mathbb{P}\{\|\mu_{t^*}-\mu_{\mathrm{SE}}\|_2\geq\epsilon'/2\}\geq\frac{\epsilon'}{2D_{\mathcal{M}}}>0$. On this event, we also know that $\lim_{\tau\to\infty}\mathrm{SR}_R(\mu_{t^*})-\mathrm{SR}_R(\mu_{\mathrm{SE}})>\delta(\epsilon'/2)$. Putting everything together, we have shown that

$$\frac{1}{a_k}\sum_{t=1}^{a_k}\lim_{\tau\to\infty}\mathbb{E}\mathrm{SR}_R(\mu_t)-\mathrm{SR}_R(\mu_{\mathrm{SE}})\geq\Delta>0,$$

and this holds for all terms in the sequence $\{a_k\}$. This finally implies that $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\mathrm{SR}_R(\mu_t)-\mathrm{SR}_R(\mu_{\mathrm{SE}})\not\to 0$. Since this contradicts the hypothesis, we conclude that $\lim_{T\to\infty}\lim_{\tau\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\|\mu_t-\mu_{\mathrm{SE}}\|_2=0$. □

A.1 Proof of Theorem 3.1

We let $\widehat{\mathrm{SR}}_L(\theta) = \mathbb{E}_{v\sim\mathrm{Unif}(\mathcal{B})}[\mathrm{SR}_L(\theta+\delta v)]$, where $\mathcal{B}$ denotes the unit ball. Then, we know that

$$\nabla\widehat{\mathrm{SR}}_L(\theta) = \frac{d}{\delta}\mathbb{E}_{u\sim\mathcal{S}}[\mathrm{SR}_L(\theta+\delta u)u],$$

where $\mathcal{S}$ denotes the unit sphere. Denote by $\hat{\theta}_{\mathrm{SE}}$ the optimum of $\widehat{\mathrm{SR}}_L$, and notice that $\widehat{\mathrm{SR}}_L$ is convex since SR_L is convex.

For any fixed t, we have

$$\begin{aligned}
\|\phi_{t+1}-\hat{\theta}_{\mathrm{SE}}\|_2^2 &\le \|\phi_t - \eta_t\frac{d}{\delta}L(\bar{\mu}_t,\phi_t+\delta u_t)u_t - \hat{\theta}_{\mathrm{SE}}\|_2^2 \\
&\le \|\phi_t-\hat{\theta}_{\mathrm{SE}}\|_2^2 - 2\eta_t\frac{d}{\delta}L(\bar{\mu}_t,\phi_t+\delta u_t)u_t^\top(\phi_t-\hat{\theta}_{\mathrm{SE}}) + \eta_t^2\frac{d^2}{\delta^2}\|L(\bar{\mu}_t,\phi_t+\delta u_t)u_t\|_2^2 \\
&\le \|\phi_t-\hat{\theta}_{\mathrm{SE}}\|_2^2 - 2\eta_t\frac{d}{\delta}L(\bar{\mu}_t,\phi_t+\delta u_t)u_t^\top(\phi_t-\hat{\theta}_{\mathrm{SE}}) + \eta_t^2\frac{d^2B^2}{\delta^2}. \qquad (6)
\end{aligned}$$

Focusing on the middle term, we have

$$\begin{aligned}
L(\bar{\mu}_t,\phi_t+\delta u_t)u_t^\top(\phi_t-\hat{\theta}_{\mathrm{SE}}) &= L(\bar{\mu}_t,\phi_t+\delta u_t)u_t^\top(\phi_t-\hat{\theta}_{\mathrm{SE}}) \pm L(\mu_{\mathrm{BR}}(\theta_t),\phi_t+\delta u_t)u_t^\top(\phi_t-\hat{\theta}_{\mathrm{SE}}) \\
&\ge L(\mu_{\mathrm{BR}}(\phi_t+\delta u_t),\phi_t+\delta u_t)u_t^\top(\phi_t-\hat{\theta}_{\mathrm{SE}}) - \beta_\mu\|\bar{\mu}_t-\mu_{\mathrm{BR}}(\theta_t)\|_2 D_\Theta.
\end{aligned}$$

Denote $\epsilon_t \overset{\mathrm{def}}{=} \mathbb{E}\|\bar{\mu}_t-\mu_{\mathrm{BR}}(\theta_t)\|_2$. Taking expectations of both sides, we get

$$\mathbb{E}L(\bar{\mu}_t,\phi_t+\delta u_t)u_t^\top(\phi_t-\hat{\theta}_{\mathrm{SE}}) \ge L(\mu_{\mathrm{BR}}(\phi_t+\delta u_t),\phi_t+\delta u_t)u_t^\top(\phi_t-\hat{\theta}_{\mathrm{SE}}) - \beta_\mu D_\Theta\epsilon_t.$$

Returning to equation (6) and taking expectations of both sides, we get

$$\begin{aligned}
\mathbb{E}\|\phi_{t+1}-\hat{\theta}_{\mathrm{SE}}\|_2^2 &\le \mathbb{E}\|\phi_t-\hat{\theta}_{\mathrm{SE}}\|_2^2 - 2\eta_t(\mathbb{E}[\nabla\widehat{\mathrm{SR}}_L(\phi_t)^\top(\phi_t-\hat{\theta}_{\mathrm{SE}})] - \beta_\mu D_\Theta\epsilon_t) + \eta_t^2\frac{d^2B^2}{\delta^2} \\
&\le \mathbb{E}\|\phi_t-\hat{\theta}_{\mathrm{SE}}\|_2^2 - 2\eta_t(\mathbb{E}\widehat{\mathrm{SR}}_L(\phi_t) - \widehat{\mathrm{SR}}_L(\hat{\theta}_{\mathrm{SE}}) - \beta_\mu D_\Theta\epsilon_t) + \eta_t^2\frac{d^2B^2}{\delta^2},
\end{aligned}$$

where in the last line we use the fact that $\widehat{\mathrm{SR}}_L$ is convex. After rearranging, we have

$$\mathbb{E}\widehat{\mathrm{SR}}_L(\phi_t) - \widehat{\mathrm{SR}}_L(\hat{\theta}_{\mathrm{SE}}) \le \frac{1}{2\eta_t}\left(\mathbb{E}\|\phi_t-\hat{\theta}_{\mathrm{SE}}\|_2^2 - \mathbb{E}\|\phi_{t+1}-\hat{\theta}_{\mathrm{SE}}\|_2^2\right) + \frac{\eta_t d^2B^2}{2\delta^2} + \beta_\mu D_\Theta\epsilon_t.$$

Summing up over $t\in\{1,\dots,T\}$, we get

$$\begin{aligned}
\sum_{t=1}^T(\mathbb{E}[\widehat{\mathrm{SR}}_L(\phi_t)] - \widehat{\mathrm{SR}}_L(\hat{\theta}_{\mathrm{SE}})) &\le \frac{D_\Theta^2}{2\eta_1} + \frac{1}{2}\sum_{t=1}^{T-1}\left(\frac{1}{\eta_{t+1}}-\frac{1}{\eta_t}\right)D_\Theta^2 + \frac{d^2B^2}{2\delta^2}\sum_{t=1}^T\eta_t + \beta_\mu D_\Theta\sum_{t=1}^T\epsilon_t \\
&\le \frac{D_\Theta^2}{2\eta_T} + \frac{d^2B^2}{2\delta^2}\sum_{t=1}^T\eta_t + \beta_\mu D_\Theta\sum_{t=1}^T\epsilon_t,
\end{aligned}$$

where we use the fact that η_t is non-increasing.

We use the fact that SR_L is Lipschitz to bound the difference between SR_L and $\widehat{\mathrm{SR}}_L$:

$$\left|\mathbb{E}[\widehat{\mathrm{SR}}_L(\phi_t) - \mathrm{SR}_L(\theta_t)]\right| \le 2\beta\delta,$$

and similarly

$$\min_{\theta\in\Theta}(\widehat{\mathrm{SR}}_L(\theta) - \mathrm{SR}_L(\theta) + \mathrm{SR}_L(\theta)) \ge \min_\theta \mathrm{SR}_L(\theta) - \beta\delta.$$

Putting everything together, we conclude

$$\sum_{t=1}^{T}(\mathbb{E}[\mathrm{SR}_L(\theta_t)] - \mathrm{SR}_L(\theta_{\mathrm{SE}})) \le \frac{D_\Theta^2}{2\eta_T} + \frac{d^2B^2}{2\delta^2}\sum_{t=1}^{T}\eta_t + 3\beta\delta T + \beta_\mu D_\Theta \sum_{t=1}^{T}\epsilon_t.$$

Setting $\eta_t = \eta_0 d^{-\frac{1}{2}} t^{-\frac{3}{4}}$ and $\delta = \delta_0 d^{\frac{1}{2}} T^{-1/4}$ yields the final bound:

$$\sum_{t=1}^{T}(\mathbb{E}[\mathrm{SR}_L(\theta_t)] - \mathrm{SR}_L(\theta_{\mathrm{SE}})) \le \left(\frac{D_\Theta^2}{2\eta_0} + \frac{2B^2}{\delta_0^2}\right)\sqrt{d}T^{3/4} + \beta_\mu D_\Theta \sum_{t=1}^{T}\epsilon_t.$$

For the second statement, observe that

$$\|\bar{\mu}_t - \mu_{\mathrm{BR}}(\theta_t)\|_2 \le \frac{1}{\tau}\sum_{j=1}^{\tau}\|\mu_{t,j} - \mu_{\mathrm{BR}}(\theta)\|_2,$$

and the right-hand side tends to zero in expectation as $\tau \to \infty$ by Lemma A.1.

A.2 Proof of Theorem 3.3

By standard convergence guarantees of gradient descent on PL objectives [26], we have

$$\|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\|_2 \le \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2}\|\mu_{t-1} - \mu_{\mathrm{BR}}(\theta_t)\|_2,$$

where $\kappa \stackrel{\mathrm{def}}{=} \frac{\beta_\mu^R}{\gamma}$. Denote $\epsilon_t \stackrel{\mathrm{def}}{=} \|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\|_2$. We will show that ϵ_t decays fast enough due to the decay in η_t. In particular, we have

$$\begin{aligned}
\epsilon_t = \|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\|_2 &\le \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2}\|\mu_{t-1} - \mu_{\mathrm{BR}}(\theta_t)\|_2 \\
&= \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2}\|\mu_{t-1} - \mu_{\mathrm{BR}}(\theta_{t-1}) + \mu_{\mathrm{BR}}(\theta_{t-1}) - \mu_{\mathrm{BR}}(\theta_t)\|_2 \\
&\le \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2}\left(\|\mu_{t-1} - \mu_{\mathrm{BR}}(\theta_{t-1})\|_2 + \|\mu_{\mathrm{BR}}(\theta_{t-1}) - \mu_{\mathrm{BR}}(\theta_t)\|_2\right) \\
&\le \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2}\|\mu_{t-1} - \mu_{\mathrm{BR}}(\theta_{t-1})\|_2 \\
&\quad + \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2}\frac{\eta_t d\beta_{\mathrm{BR}}}{\delta}\|L(\mu_t, \phi_t + \delta u_t)u_t\|_2 \\
&\le \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2}\epsilon_{t-1} + \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2}\frac{\eta_t d\beta_{\mathrm{BR}}}{\delta}B.
\end{aligned}$$

Now suppose τ is chosen such that $\tau > \frac{\log(\kappa)}{\log\left(\frac{1}{1-\gamma\eta_\mu}\right)}$. Then we have that $\alpha(\tau) \stackrel{\mathrm{def}}{=} \sqrt{\kappa}(1-\gamma\eta_\mu)^{\tau/2} < 1$. (Note that as τ increases, $\alpha(\tau)$ can be driven to zero.) Altogether, we find that:

$$\epsilon_t \le \alpha(\tau)\epsilon_{t-1} + \alpha(\tau)\eta_t\frac{d\beta_{\mathrm{BR}}B}{\delta}.$$

Unrolling the recursion, we find that

$$\epsilon_t \le \alpha(\tau)^t\epsilon_0 + \frac{d\beta_{\mathrm{BR}}B}{\delta}\sum_{i=1}^{t}\alpha(\tau)^{t+1-i}\eta_i.$$

Summing up over $t \in \{1, \dots, T\}$, we get

$$\begin{aligned}
\sum_{t=1}^{T} \epsilon_t &\le \epsilon_0 \sum_{t=1}^{T} \alpha(\tau)^t + \frac{d\beta_{\mathrm{BR}} B}{\delta} \sum_{t=1}^{T} \sum_{i=1}^{t} \alpha(\tau)^{t+1-i} \eta_i \\
&\le \frac{\epsilon_0}{1-\alpha(\tau)} + \frac{d\beta_{\mathrm{BR}} B}{\delta} \sum_{t=1}^{T} \sum_{i=1}^{T} \alpha(\tau)^{t+1-i} \eta_i \mathbf{1}\{i \le t\} \\
&= \frac{\epsilon_0}{1-\alpha(\tau)} + \frac{d\beta_{\mathrm{BR}} B}{\delta} \sum_{i=1}^{T} \eta_i \sum_{t=1}^{T} \alpha(\tau)^{t+1-i} \mathbf{1}\{i \le t\} \\
&= \frac{\epsilon_0}{1-\alpha(\tau)} + \frac{d\beta_{\mathrm{BR}} B}{\delta} \sum_{i=1}^{T} \eta_i \sum_{t=i}^{T} \alpha(\tau)^{t+1-i} \\
&\le \frac{\epsilon_0}{1-\alpha(\tau)} + \frac{d\beta_{\mathrm{BR}} B}{\delta(1-\alpha(\tau))} \sum_{t=1}^{T} \eta_t.
\end{aligned}$$

For $\eta_t = \eta_0 d^{-1/2} t^{-3/4}$ and $\delta = \delta_0 d^{1/2} T^{-1/4}$, we have

$$\sum_{t=1}^{T} \epsilon_t \le \frac{1}{1-\alpha(\tau)} \left(\epsilon_0 + \frac{4\beta_{\mathrm{BR}} B \eta_0 \sqrt{T}}{\delta_0} \right).$$

A.3 Proof of Theorem 3.5

Define $\mu_t^* = D_t(\mu_1, \theta_{\mathrm{BR}}(\mu_1), \dots, \mu_{t-1}^*, \theta_{\mathrm{BR}}(\mu_{t-1}), \xi_t)$. First we will prove that

$$\lim_{T\to\infty} \lim_{\tau\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\mathrm{SR}_R(\mu_t) - \mathrm{SR}_R(\mu_{\mathrm{SE}}) = 0. \tag{7}$$

To show this, it suffices to prove that for all t, $\mu_t \to_p \mu_t^*$ as $\tau \to \infty$. The sufficiency of this condition follows because

$$\begin{aligned}
&\lim_{T\to\infty} \lim_{\tau\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\mathrm{SR}_R(\mu_t) - \mathrm{SR}_R(\mu_{\mathrm{SE}}) \\
&= \lim_{T\to\infty} \lim_{\tau\to\infty} \frac{1}{T} \sum_{t=1}^{T} [\mathbb{E}\mathrm{SR}_R(\mu_t) - \mathbb{E}\mathrm{SR}_R(\mu_t^*) + \mathbb{E}\mathrm{SR}_R(\mu_t^*)] - \mathrm{SR}_R(\mu_{\mathrm{SE}}) \\
&= \lim_{T\to\infty} \lim_{\tau\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left(\mathbb{E}\mathrm{SR}_R(\mu_t) - \mathbb{E}\mathrm{SR}_R(\mu_t^*) \right),
\end{aligned}$$

where the last step follows by the assumption that the agents play a rational strategy. Therefore, if $\mu_t \to_p \mu_t^*$, continuity of $\mathrm{SR}_R(\mu)$ implies $\mathbb{E}\mathrm{SR}_R(\mu_t) - \mathbb{E}\mathrm{SR}_R(\mu_t^*) \to 0$ and we get the desired conclusion.

We prove that $\mu_t \to_p \mu_t^*$ by induction. Notice that $\mu_1 \equiv \mu_1^*$ by definition. Suppose that $\mu_j \to_p \mu_j^*$ for all $j < t$. Denote by $\theta_{j,\tau}$ the possibly randomized algorithm that maps μ_j to θ_j. Then, for any $\mu \in \mathcal{M}$, we know that $\|\theta_{j,\tau}(\mu) - \theta_{\mathrm{BR}}(\mu)\|_2 \to_p 0$ by assumption. This in turn implies that for all $j < t$,

$$\|\theta_{j,\tau}(\mu_j) - \theta_{\mathrm{BR}}(\mu_j^*)\|_2 \le \|\theta_{j,\tau}(\mu_j) - \theta_{\mathrm{BR}}(\mu_j)\|_2 + \|\theta_{\mathrm{BR}}(\mu_j) - \theta_{\mathrm{BR}}(\mu_j^*)\|_2 \to_p 0,$$

where the second term tends to zero by the continuous mapping theorem. Finally, we can apply the continuity of D_t to conclude that $\mu_t \to_p \mu_t^*$, as desired.

Let β denote the Lipschitz constant of SR_L. Finally, we we can apply this Lipschitz condition to conclude:

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}L(\mu_t,\theta_{t,\tau}) - L(\mu_{\mathrm{SE}},\theta_{\mathrm{BR}}(\mu_{\mathrm{SE}}))\\
&= \frac{1}{T}\sum_{t=1}^{T}[\mathbb{E}L(\mu_t,\theta_{t,\tau}) \pm \mathbb{E}L(\mu_t,\theta_{\mathrm{BR}}(\mu_t)))] - L(\mu_{\mathrm{SE}},\theta_{\mathrm{BR}}(\mu_{\mathrm{SE}}))\\
&= \frac{1}{T}\sum_{t=1}^{T}(\mathbb{E}L(\mu_t,\theta_{\mathrm{BR}}(\mu_t)) - L(\mu_{\mathrm{SE}},\theta_{\mathrm{BR}}(\mu_{\mathrm{SE}})) + \mathbb{E}[L(\mu_t,\theta_{t,\tau}) - L(\mu_t,\theta_{\mathrm{BR}}(\mu_t))])\\
&\leq \frac{\beta}{T}\sum_{t=1}^{T}\mathbb{E}\|\mu_t - \mu_{\mathrm{SE}}\|_2 + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[L(\mu_t,\theta_{t,\tau}) - L(\mu_t,\theta_{\mathrm{BR}}(\mu_t)].
\end{aligned}$$

By Lemma A.1, the guarantee (7) implies that the first term vanishes. The second term vanishes by continuity. Therefore, taking the limit over T,τ, we obtain

$$\lim_{T\to\infty}\lim_{\tau\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}L(\mu_t,\theta_t) - L(\mu_{\mathrm{SE}},\theta_{\mathrm{BR}}(\mu_{\mathrm{SE}})) = 0,$$

as desired.

A.4 Proof of Propositon 4.1

First we assume the decision-maker leads. When θ is the deployed model, the best response by the agents is to simply move by distance B in the direction of θ. Thus, $\mu_{\mathrm{BR}}(\theta)$ is given by:

$$\mu_{\mathrm{BR}}(\theta) = \arg\min_{\mu}\mathbb{E}_{(x,y)\sim\mathcal{P}(\mu)} - x^\top\theta = \frac{\theta}{\|\theta\|_2}B.$$

This implies the following expected loss for the decision-maker:

$$\begin{aligned}
L(\mu_{\mathrm{BR}}(\theta),\theta) &= \mathbb{E}_{z\sim\mathcal{P}\left(\frac{\theta}{\|\theta\|_2}B\right)}\ell(z;\theta) = \frac{1}{2}\mathbb{E}_{(x_0,y)\sim\mathcal{P}(0)}\left(y - x_0^\top\theta - \|\theta\|_2 B\right)^2\\
&= \frac{\sigma^2}{2} + \frac{1}{2}\|\beta-\theta\|_2^2 + \frac{B^2}{2}\|\theta\|_2^2.
\end{aligned}$$

This objective is convex and thus by finding a stationary point we observe that it is minimized at $\theta_{\mathrm{SE}} = \frac{\beta}{1+B^2}$. By plugging this choice back into the previous equation, we observe that the minimal Stackelberg risk of the decision-maker is equal to

$$\mathrm{SR}_L(\theta_{\mathrm{SE}}) = L(\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}),\theta_{\mathrm{SE}}) = \frac{\sigma^2}{2} + \frac{\|\beta\|_2^2 B^2}{2(1+B^2)}. \tag{8}$$

Moreover, the agents' loss at θ_{SE} is equal to:

$$R(\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}),\theta_{\mathrm{SE}}) = -\|\theta_{\mathrm{SE}}\|_2 B = -\frac{\|\beta\|_2 B}{1+B^2}.$$

Now we reverse the order of play and assume that the agents lead. If the agents move by μ; i.e., they follow the law $\mathcal{P}(\mu)$, then the decision-maker incurs loss:

$$L(\mu,\theta) = \mathbb{E}_{(x,y)\sim\mathcal{P}(\mu)}\frac{1}{2}\left(y - x_0^\top\theta - \mu^\top\theta\right)^2 = \frac{\sigma^2}{2} + \frac{1}{2}\|\beta-\theta\|_2^2 + \frac{1}{2}(\mu^\top\theta)^2.$$

By computing a stationary point, we find that the best response of the decision-maker is:

$$\theta_{\mathrm{BR}}(\mu) = (I + \mu\mu^\top)^{-1}\beta = \left(I - \frac{\mu\mu^\top}{1+\|\mu\|_2^2}\right)\beta.$$

The Stackelberg risk of the strategic agent is then

$$\begin{aligned}\mathrm{SR}_R(\mu) = R(\mu, \theta_{\mathrm{BR}}(\mu)) &= -\mu^\top \theta_{\mathrm{BR}}(\mu) = -\mu^\top \left(I - \frac{\mu\mu^\top}{1+\|\mu\|_2^2}\right)\beta \\ &= -\mu^\top\beta + \frac{\|\mu\|_2^2\mu^\top\beta}{1+\|\mu\|_2^2} = -\frac{\mu^\top\beta}{1+\|\mu\|_2^2}.\end{aligned}$$

Among all μ such that $\|\mu\|_2 = C$, $\mathrm{SR}_R(\mu)$ is minimized when μ points in the β direction: $\mu = C\frac{\beta}{\|\beta\|_2}$. With this reparameterization, we can equivalently write $\min_\mu \mathrm{SR}_R(\mu)$ as

$$\min_{C>0} \|\beta\|_2 \frac{-C}{1+C^2}.$$

This function is decreasing for $C \in (0,1]$, and increasing for $C > 1$. Therefore, $\mu_{\mathrm{SE}} = \min(1,B)\frac{\beta}{\|\beta\|_2}$, and

$$\mathrm{SR}_R(\mu_{\mathrm{SE}}) = -\|\beta\|_2 \frac{\min(1,B)}{1+\min(1,B)^2}.$$

Finally, we evaluate the decision-maker's loss at μ_{SE}:

$$\begin{aligned}L(\mu_{\mathrm{SE}}, \theta_{\mathrm{BR}}(\mu_{\mathrm{SE}})) &= \frac{\sigma^2}{2} + \frac{1}{2}\frac{(\beta^\top\mu_{\mathrm{SE}})^2\|\mu_{\mathrm{SE}}\|_2^2}{(1+\|\mu_{\mathrm{SE}}\|_2^2)^2} + \frac{1}{2}\left(\|\beta\|_2\frac{\min(1,B)}{1+\min(1,B)^2}\right)^2 \\ &= \frac{\sigma^2}{2} + \frac{1}{2}\frac{\|\beta\|_2^2\min(1,B)^4}{(1+\min(1,B)^2)^2} + \frac{1}{2}\left(\|\beta\|_2\frac{\min(1,B)}{1+\min(1,B)^2}\right)^2 \\ &= \frac{\sigma^2}{2} + \frac{\|\beta\|_2^2\min(1,B)^2}{2(1+\min(1,B)^2)}.\end{aligned}$$

A.5 Proof of Proposition 4.2

First we evaluate $L(\mu,\theta)$:

$$\begin{aligned}L(\mu,\theta) &= \mathbb{E}_{(x,y)\sim\mathcal{P}(\mu)}\left[-yx^\top\theta + \log(1+e^{x^\top\theta})\right] \\ &= \mathbb{E}_{(x,y)\sim\mathcal{P}(\mu)}\left[\log(e^{-yx^\top\theta} + e^{(1-y)x^\top\theta})\right] \\ &= p\mathbb{E}_{x_0\sim N(\alpha,\sigma^2 I)}\log(1+e^{-x_0^\top\theta}) + (1-p)\mathbb{E}_{x_0\sim N(\alpha,\sigma^2 I)}\log(1+e^{-x_0^\top\theta+\mu^\top\theta}) \\ &= p\mathbb{E}_{z\sim N(0,\sigma^2)}\log(1+e^{-\alpha^\top\theta+\|\theta\|_2 z}) + (1-p)\mathbb{E}_{z\sim N(0,\sigma^2)}\log(1+e^{-\alpha^\top\theta+\|\theta\|_2 z+\mu^\top\theta}).\end{aligned}$$

We prove that the agents are never worse off if they lead, for all $p \in [0,1]$. We will provide a sufficient condition; namely, we will show that

$$\mathrm{SR}_R\left(\frac{\alpha}{\|\alpha\|_2}B\right) = R(\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}), \theta_{\mathrm{SE}}).$$

This immediately implies that $\mathrm{SR}_R(\mu_{\mathrm{SE}}) \le R(\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}), \theta_{\mathrm{SE}})$.

To see this, first observe that

$$R(\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}), \theta_{\mathrm{SE}}) = B\|\theta_{\mathrm{SE}}\|_2,$$

where

$$\theta_{\mathrm{SE}} = \frac{\alpha}{\|\alpha\|_2}\cdot\min_{C_\theta}\mathbb{E}_{z\sim N(0,\sigma^2)}\left[p\log(1+e^{-\|\alpha\|_2 C_\theta + C_\theta z}) + (1-p)\log(1+e^{-\|\alpha\|_2 C_\theta + BC_\theta + C_\theta z})\right]. \tag{9}$$

Here we use the fact that the best response of the agents is to simply move by distance B in the direction of θ:

$$\mu_{\mathrm{BR}}(\theta) = \arg\max_{\mu \in \mathcal{M}} \theta^\top \mu = \frac{\theta}{\|\theta\|_2} B.$$

Now we evaluate the decision-maker's best response to $\mu = \frac{\alpha}{\|\alpha\|_2} B$. We have

$$\begin{aligned}
&\arg\min_\theta L\left(\frac{\alpha}{\|\alpha\|_2} B, \theta\right) \\
&= \arg\min_\theta \mathbb{E}_{z \sim N(0,\sigma^2)} \left[p \log(1 + e^{-\alpha^\top \theta + \|\theta\|_2 z}) + (1-p) \log(1 + e^{\alpha^\top \theta (-1 + \frac{B}{\|\alpha\|_2}) + \|\theta\|_2 z}) \right].
\end{aligned}$$

Since $B \leq \|\alpha\|_2$, we have

$$\begin{aligned}
&\arg\min_\theta L\left(\frac{\alpha}{\|\alpha\|_2} B, \theta\right) \\
&= \frac{\alpha}{\|\alpha\|_2} \cdot \min_{C_\theta} \mathbb{E}_{z \sim N(0,\sigma^2)} \left[p \log(1 + e^{-\|\alpha\|_2 C_\theta + C_\theta z}) + (1-p) \log(1 + e^{\|\alpha\|_2 C_\theta \left(-1 + \frac{B}{\|\alpha\|_2}\right) + C_\theta z}) \right].
\end{aligned}$$

Returning to equation (9), we see that $\theta_{\mathrm{SE}} = \theta_{\mathrm{BR}}\left(\frac{\alpha}{\|\alpha\|_2} B\right)$. Moreover, $\mathrm{SR}_R\left(\frac{\alpha}{\|\alpha\|_2} B\right) = B\|\theta_{\mathrm{SE}}\|_2$, as desired.

Now we analyze the decision-maker's preference. We write

$$L(\mu, \theta) = -\alpha^\top \theta + p \mathbb{E}_{x_0 \sim N(\alpha, \sigma^2 I)} \log(e^{x_0^\top \theta} + 1) + (1-p) \mathbb{E}_{x_0 \sim N(\alpha, \sigma^2 I)} \log(e^{x_0^\top \theta} + e^{\mu^\top \theta}).$$

This loss is increasing in $\mu^\top \theta$; that is, for any θ it holds that $\max_\mu L(\mu, \theta) = L(\mu_{\mathrm{BR}}(\theta), \theta)$. Using this, we observe that for every θ we have

$$L(\mu_{\mathrm{SE}}, \theta_{\mathrm{BR}}(\mu_{\mathrm{SE}})) \leq L(\mu_{\mathrm{SE}}, \theta) \leq \max_{\mu \in \mathcal{M}} L(\mu, \theta) = \mathrm{SR}_L(\theta).$$

Since this also holds for $\theta = \theta_{\mathrm{SE}}$, we conclude that following is never worse than leading for the decision-maker.

B Local guarantees when SR_L is nonconvex

We provide local guarantees in the limit when the decision-maker's Stackelberg risk is possibly nonconvex. Specifically, we show that the update rule (3) converges to a stationary point of a smooth version of the decision-maker's Stackelberg risk, provided that the strategic agents achieve iterate convergence. Moreover, under mild regularity conditions, this stationary point is a local minimum (see, e.g., Theorem 9.1 in [3]).

Proposition B.1. *Assume that the agents achieve iterate convergence, $\|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\| \to 0$ almost surely as $t \to \infty$. Further, assume that the decision-maker's Stackelberg risk is Lipschitz and smooth, and that $L(\mu, \theta)$ is Lipschitz in its first argument and bounded for all μ and θ. If the decision-maker runs update (3) with η_t satisfying $\sum_{t=1}^\infty \eta_t = \infty$ and $\sum_{t=1}^\infty \eta_t^2 < \infty$, and $\delta = \frac{\epsilon}{4\beta}$, then as $t \to \infty$, $\phi_t \to \phi^*$ such that $\nabla \widehat{\mathrm{SR}}_L(\phi^*) = 0$, where $\widehat{\mathrm{SR}}_L(\phi) = \mathbb{E}_{v \sim \mathrm{Unif}(\mathcal{B})} [\mathrm{SR}_L(\phi + \delta v)]$.*

Proof. The proof makes use of results from the literature on stochastic approximation [see, e.g., Theorem 9.1 in 6].

As in the proof of Theorem 3.1, let $\widehat{\mathrm{SR}}_L(\phi) = \mathbb{E}_{v\sim\mathrm{Unif}(\mathcal{B})}\left[\mathrm{SR}_L(\phi+\delta v)\right]$, and recall $\theta_t = \phi_t + \delta u_t$, $u_t \sim \mathrm{Unif}(\mathcal{S}^{d-1})$. We begin by writing the update for ϕ_t as:

$$\phi_{t+1} = \phi_t - \eta_t \left(\nabla_\phi \widehat{\mathrm{SR}}_L(\phi_t) - \left(\underbrace{\nabla_\phi \widehat{\mathrm{SR}}_L(\phi_t) - \frac{d}{\delta} L(\mu_{\mathrm{BR}}(\theta_t), \theta_t) u_t}_{=\mathrm{I}} \right) \right) - \eta_t \frac{d}{\delta} \left(\underbrace{L(\mu_t, \theta_t) u_t - L(\mu_{\mathrm{BR}}(\theta_t), \theta_t) u_t}_{=\mathrm{II}} \right).$$

Since

$$\nabla_\phi \widehat{\mathrm{SR}}_L(\phi) = \mathbb{E}_{u\sim\mathrm{Unif}(\mathcal{S}^{d-1})}\left[\frac{d}{\delta} \mathrm{SR}_L(\phi+\delta u) u \right],$$

we know $\mathbb{E}_u[\mathrm{I}] = 0$. Since L is bounded and SR_L is smooth, we know $\|\mathrm{I}\|_2$ is bounded. Thus, I is a zero-mean random variable with finite variance.

For term II, we use the assumed Lipshchitzness of L in the first argument to find that:

$$\begin{aligned} \|\mathrm{II}\|_2 &\le \beta_\mu \|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\|_2 \|u_t\|_2 \\ &= \beta_\mu \|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\|_2, \end{aligned}$$

where we use the fact that $\|u\|_2 = 1$. By assumption, $\|\mu_t - \mu_{\mathrm{BR}}(\theta_t)\|_2 \to 0$ almost surely as $t \to \infty$. Thus, we can write the update rule as:

$$\phi_{t+1} = \phi_t - \eta_t \left(\nabla_\phi \widehat{\mathrm{SR}}_L(\phi_t) + \xi_t + M_t \right),$$

where $\xi_t = o(1)$ and M_t is a zero-mean random variable with finite variance. Since the assumed choice of η_t satisfies $\sum_{t=1}^\infty \eta_t = \infty$ and $\sum_{t=1}^\infty \eta_t^2 < \infty$ we can invoke Chapter 2, Corollary 3 in [6] to find that $\phi_t \to \phi^* \in \{\phi : \nabla_\phi \widehat{\mathrm{SR}}_L(\phi) = 0\}$. □

C Experimental details

To generate Figure 1, we first find the decision-maker's equilibrium by optimizing $L\left(\frac{\theta}{\|\theta\|_2}B, \theta\right)$, as given in the proof of Proposition 4.2, with 1000 steps of gradient descent. We approximate the relevant expectation via a sample average over 1000 samples $x_0 \sim N(\alpha, \sigma^2)$. Once we have θ_{SE}, we compute $\mu_{\mathrm{BR}}(\theta_{\mathrm{SE}}) = \frac{\theta_{\mathrm{SE}}}{\|\theta_{\mathrm{SE}}\|_2}B$. To find the agents' equilibrium, we perform grid search over 1000 equally spaced points in the interval $[-B, B]$. For each point in the grid, we compute the relevant best response $\theta_{\mathrm{BR}}(\mu)$ by running 1000 steps of gradient descent, again estimating the expectation over 1000 samples. We take as the agents' equilibrium the point μ in the grid that minimizes the estimated value of $R(\mu, \theta_{\mathrm{BR}}(\mu))$.

To generate Figures 2 and 3, in all experiments we set the step size of the faster player, that is, the one running gradient descent, to 0.1. When the decision-maker leads, we set their step size to be $\eta_t = \eta_0 t^{-3/4}$, where $\eta_0 = 6$ for $p = 0.1$, $\eta_0 = 5$ for $p = 0.5$, and $\eta_0 = 10$ for $p = 0.9$. When the agents lead, we set their step size to be $\eta_t = 0.02t^{-3/4}$. We let the perturbation parameter δ decrease with time, and set $\delta_t = t^{-1/4}$.

To generate Figure 4, when the decision-maker leads, we set their step size to be $\eta_t = 0.1t^{-3/4}$. When the agents lead, we set their step size to be $\eta_t = 0.01t^{-3/4}$. We let the perturbation parameter δ decrease with time, and set $\delta_t = t^{-1/4}$. When the decision-maker and agents follow, their step sizes for gradient descent are 0.1 and 0.01, respectively. To generate Figure 5 all parameters are kept the same except when the decision-maker leads, we set their step size to be $\eta_t = t^{-3/4}$.

To compute the decision-maker's and agents' risk at the decision-maker's equilibrium, we explicitly compute the decision-maker's Stackelberg risk by using the fact that the best response of the agents is

$\mu_{\mathrm{BR}}(\theta) = \frac{\theta}{\|\theta\|_2} B$ and $\mu_{\mathrm{BR}}(\theta) = \frac{1}{\lambda}\theta$ in the two respective settings. We then minimize this risk directly using gradient descent with a fixed step size of 0.1.

To compute the decision-maker's and agents' risk at the agents' equilibrium, we compute their Stackelberg risk over a grid by fixing μ and running gradient descent with step size 0.1 on the decision-maker's problem until convergence for each value of μ. We then find the agents' equilibrium by searching for the minimal estimate of $R(\mu, \theta_{\mathrm{BR}}(\mu))$ over the grid.