Performative Power

Anonymous Author(s)
Affiliation
Address
email

Abstract

We introduce the notion of *performative power*, which measures the ability of a firm operating an algorithmic system, such as a digital content recommendation platform, to steer a population. We relate performative power to the economic theory of market power. Traditional economic concepts are well known to struggle with identifying anti-competitive patterns in digital platforms—a core challenge is the difficulty of defining the market, its participants, products, and prices. Performative power sidesteps the problem of market definition by focusing on a directly observable statistical measure instead. High performative power enables a platform to profit from *steering* participant behavior, whereas low performative power ensures that *learning* from historical data is close to optimal. Our first general result shows that under low performative power, a firm cannot do better than standard supervised learning on observed data. We draw an analogy with a firm being a price-taker, an economic condition that arises under perfect competition in classical market models. We then contrast this with a market where performative power is concentrated and show that the equilibrium state can differ significantly. We go on to study performative power in a concrete setting of strategic classification where participants can switch between competing firms. We show that monopolies maximize performative power and disutility for the participant, while competition and outside options decrease performative power. We end on a discussion of connections to measures of market power in economics and of the relationship with ongoing antitrust debates.

1 Introduction

Digital platforms pose a well-recognized challenge for antitrust enforcement. Traditional market definitions, along with associated notions of competition and market power, map poorly onto digital platforms. An authoritative report, published by the Stigler Committee [2019], details the many challenges associated with digital platforms, among them: *"Pinpointing the locus of competition can also be challenging because the markets are multisided and often ones with which economists and lawyers have little experience. This complexity can make market definition another hurdle to effective enforcement."* Published the same year, a comprehensive report from the European Commission calls for *"less emphasis on analysis of market definition, and more emphasis on theories of harm and identification of anti-competitive strategies."* [Crémer et al., 2019]

Our work responds to this call by developing a normative and technical proposal for reasoning about power in digital economies, while sidestepping market definition. Our running example is a digital content recommendation platform. The platform connects content creators with viewers, while monetizing views through digital advertisement. Key to the business strategy of a firm operating a digital content recommendation platform is its ability to predict revenue for content that it recommends or ranks highly. Often framed as a supervised learning task, the firm trains a statistical model on observed data to predict some proxy of revenue, such as clicks, views, or engagement. Better predictions increase profit for the firm.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

A second way of increasing profit is more subtle. The platform can use its predictions to steer participants towards modes of consumption and production that are easier to predict and monetize. For example, the platform could reward consistency in the videos created by content creators, so that the audience and the popularity of their videos becomes more predictable. Similarly, the platform could promote content that shapes viewer preferences, so that viewer interactions are easier to monetize. How potent this strategy is depends on the extent to which the firm is able to steer participants, which we argue is an important indicator of the power of the firm.

1.1 Our contribution

We introduce the notion of *performative power* that quantifies a firm's ability to steer a population of participants. We build on recent developments in performative prediction [Perdomo et al., 2020] to formalize the fundamental difference between learning and steering. Performativity allows the predictor to influence the data-generating process, a dependence ruled out by the traditional theory of supervised learning. As such performative power provides an additional lever a firm can exploit towards maximizing their objective function, either explicitly or implicitly.

Performative power is a causal statistical notion that directly quantifies how much participants change in response to updates to the predictive model operated by the platform. In doing so it avoids market specifics, such as the number of firms involved, products, and monetary prices. Neither does it require a competitive equilibrium notion as a reference point. Instead, it focuses on where rubber meets the road: the predictive model of the platform and its causal powers. We argue that the sensitivity of participant behavior to algorithmic changes in the platform reveals a salient power relationship between the platform and the participant.

Our first general result shows that under low performative power, a firm cannot do better than standard supervised learning on observed data. Intuitively, this means the firm optimizes its loss function *ex-ante* on data it observes without the ability to steer towards data it would prefer. We interpret this optimization strategy as analogous to the firm being a price-taker, an economic condition that arises under perfect competition in classical market models. We contrast this optimization strategy with a firm that performs *ex-post* optimization and takes advantage of the performative power it may have to achieve their goals.

To better understand the implications of our definition we study performative power in the concrete market model of strategic classification. Strategic classification instantiates performative prediction with standard microfoundations by which participants change their features rationally in response to a predictor. In this simple setting, the strength of these strategic effects governs the performative power of the firm. We study the effects of different economic factors by extending the standard model to incorporate competing firms and outside options. We highlight three observations:

i) A *monopoly* firm maximizes performative power. In this case, participants are willing to incur a cost up to the utility of using the service in order to adjust to the firm's predictor.

ii) Performative power decreases in the presence of *outside options*.

iii) When there is *competition* for participants and firms offer services that are perfect substitutes for each other, then even two firms can lead to zero performative power. This result stands in analogy with the classical Bertrand competition.

We end on a discussion of the relationship between performative power and measures of market power in economics and of the role of performative power in ongoing antitrust debates.

1.2 Related Work

The phenomenon of performativity in the context of prediction has been discussed and formalized by Perdomo et al. [2020]. Picking up on the framework of *performative prediction*, optimization challenges and solution concepts in the presence of performativity have been studied in various works, including [Perdomo et al., 2020; Mendler-Dünner et al., 2020; Izzo et al., 2021; Dong and Ratliff, 2021; Miller et al., 2021; Brown et al., 2020; Li and Wai, 2021; Ray et al., 2022; Jagadeesan et al., 2022; Wood et al., 2022]. Our work provides an complementary perspective in that we focus on the role of *steering*, rather than the convergence of learning algorithms to solutions of low risk.

The *strategic classification* setup we use for our case study to instantiate performative prediction with a specific microfoundation model was proposed in [Brückner et al., 2012; Hardt et al., 2016]. The standard model has received significant attention from both the machine learning community [Milli et al., 2019; Hu et al., 2019; Braverman and Garg, 2020; Kleinberg and Raghavan, 2019; Dong et al., 2018; Zrnic et al., 2021; Levanon and Rosenfeld, 2021] and the economics community [Frankel and Kartik, 2020; Ball, 2020; Hennessy and Goodhart, 2020; Frankel and Kartik, 2019]. While the perfect rationality and information assumptions underlying the standard model have been questioned in various works [Jagadeesan et al., 2021; Ghalme et al., 2021; Bechavod et al., 2021], the focus has exclusively been on describing participant behavior in response to a single firm acting in isolation. Similarly, in performative prediction more broadly, existing works have predominantly focused on a single decision-maker, apart from Narang et al. [2022] and Piliouras and Yu [2022].

2 Performative power

Fix a set $\mathcal{U}$ of participants, each associated with a data point $z(u)$. Fix a metric $\mathrm{dist}(z, z')$ over the space of data points. We assume that participants interact with a designated firm. At a fixed point in time, the firm can take actions in a set $\mathcal{F}$. We think of an action $f \in \mathcal{F}$ as a predictor that the firm can deploy at that point in time. We denote by $z_f(u)$ a potential outcome random variable that specifies the data point for participant u that were to result from the firm's hypothetical action f.

Definition 1 (Performative Power)**.** *Let $\mathcal{U}$ be a population of participants interacting with a firm. Denote by $\mathcal{F}$ a set of actions the firm can take. For each participant $u \in \mathcal{U}$ and action $f \in \mathcal{F}$, let $z(u)$ be the data of participant u and let $z_f(u)$ be a potential outcome variable representing the data of participant u if the firm were to take action f. We define the* performative power *of the firm with respect to the population, action set, and potential outcomes as*

$$\mathrm{P} := \sup_{f \in \mathcal{F}} \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}\left[\mathrm{dist}\left(z(u), z_f(u)\right)\right] .$$

Here, $\mathrm{dist}(z, z')$ is a metric over the space of data points. The expectation is over the randomness in the potential outcomes.

The expression inside the supremum generalizes an average treatment effect, corresponding to scalar valued potential outcomes and the absolute value as metric. We could generalize other causal quantities such as heterogeneous treatment effects, but this avenue is not subject of our paper. The definition takes a supremum over possible actions a firm can take at a specific point in time. We can therefore lower bound performative power by estimating the causal effect of any given action $f \in \mathcal{F}$. In practice, the action set $\mathcal{F}$ we are interested in typically corresponds to predictive models that increase profit for the firm. The choice of the population $\mathcal{U}$ depends on what power relationship we would like to investigate.

Having specified $\mathcal{F}$ and $\mathcal{U}$, estimating performative power amounts to causal inference involving the potential outcome variables $z_f(u)$ for unit $u \in \mathcal{U}$ and action $f \in \mathcal{F}$. In an observational design, a researcher might observe participant data shortly before and after a platform updated a model. In an experimental design, the researcher would deploy a suitable chosen predictor to estimate the effect. Neither route requires understanding the market in which the firm operates. It is not even necessary to know the firm's objective function, how it optimizes its objective, and whether it successfully achieves its objective. In practice, the dynamic process that generates the potential outcome $z_f(u)$ may be highly complex, but this complexity does not enter the definition. Consequently, the definition applies to complex multisided digital economies that resist a clean definition.

Display discontinuity design. At first glance, it may appear that computing performative power requires access to participant data at multiple time steps. However, we propose a method—that we call that *display discontinuity design*—which only relies on historical data from a *single* time snapshot to measure performative power in a rich class of examples. Consider a digital content recommendation system where the platform recommends content to each user in several display slots. The ordering of content across display slots is well-known to shape consumption patterns: in particular, methods based on regression discontinuity can be applied to measure the magnitude of these "position biases" from a single time snapshot Narayanan and Kalyanam [2015]. We translate these position bias estimates to a lower bound on performative power by constructing a decision rule

f^* from the current decision rule f that swaps the content in two adjacent slots (e.g. the first two slots) for each user, but leaves all other recommendations unchanged. The difference in position bias between the two slots can be used to lower bound an appropriate instantation of performative power. We defer the details of this example to Appendix A. We hope that this example serves as a proof of concept for using display discontinuity design to compute performative power.

2.1 Running example: Digital content recommendation

To work out our running example in greater detail, think of a digital content recommendation platform, such as the video sharing services YouTube or Twitch. The platform aims to recommend channels that generate a high revenue, and these recommendations are typically personalized to each viewer. Towards this goal, it is helpful for the platform to collect data and build a predictor f for predicting the value of recommending a channel c to a viewer with preferences p. Let $x = (x_c, x_p)$ be the features used for the prediction task that capture attributes x_c of the channel and the attributes x_p of the viewer preferences. Let y be the target variable, such as *watch time*, that acts as a proxy for the monetary value of showing a channel to a specific viewer. For concreteness, take the supervised learning loss $\ell(f(x), y)$ incurred by a predictor f to be the squared loss $(f(x) - y)^2$.

When defining performative power, participants could be viewers or content creators. The definition applies to both, and we show how to instantiate performative power in each case. By selecting the units $\mathcal{U}$, which features to include in the data point z, and how to specify the distance metric dist, we can pinpoint the power relationship we would like to investigate.

Content creators. For content creators, the predictor f can affect the type of videos that they stream on their channels. For example, content creators might strategically adjust various features of their content relevant for the predicted outcome, such as the length, type or description of their videos, to improve their ranking. To investigate the effect of the predictor on the type of content being produced, let the participants $\mathcal{U}$ be content creators and suppose that each content creator $u \in \mathcal{U}$ maintains a channel of videos. Let the data point z correspond to features x_c characterizing the channel c created by content creator u. The resulting instantiation of performative power measures the changes in content with respect to dist induced by the firm actions. (See Appendix B for a simple, model for content creator behavior that can induce changes in content.)

Viewers. For viewers, the predictor f can shape the preferences of viewers, potentially steering these preferences towards recommended content. (Note that this is a different type of performative effect than the one measured via display discontinuity design.) To investigate the effect of the predictor on viewer preferences, think of the participants $\mathcal{U}$ as viewers, and let the data point z correspond to attributes x_p characterizing the preferences of viewer u. The resulting instantiation of performative power measures the changes in preferences with respect to dist induced by the firm actions. (See Appendix B for a simple model for viewer behavior that induces preference changes.)

To investigate the role of performativity in the firm's optimization problem, we focus on the simplified task of predicting the watch time for each channel on the platform for a given viewer with preferences x_p.[1] The units in this optimization problem are content producers u that own a channel with attributes $x(u)$. The associated label $y(u)$ corresponds to the watch time for the viewer on that channel. Since performative effects can occur both in content creation and viewer preferences, both x and y could change with the predictor f.

In this setup, we distinguish between the *ex-ante* loss $\ell(f(x(u)), y(u))$ and the *ex-post* loss $\ell(f(x_f(u)), y_f(u))$. The former describes the loss that the firm can optimize when building the predictor. The latter describes the loss that the firm observes after deploying f. Platforms typically build predictive models by first optimizing over historical data and then A/B test candidate models to check their ex-post effectiveness (we refer to [Covington et al., 2016] for some insights on the Youtube recommendation pipeline). The firm thus implicitly performs some degree of ex-post optimization.[2]

[1] If we instead wished to focus on viewers, we could instead pose the optimization problem as predicting the watch time for each viewer for a single content producer's channel. The units in this case would be the viewers.

[2] While the firm ultimately cares about profit, we note that the learning problem inevitably optimizes some surrogate objective, such as clicks, watch time, likes or other metrics. We discuss this in Appendix B.2.

3 Learning versus steering

High performative power represents a reason for concern because it offers a lever that the firm can exploit towards achieving their goal. Instead of identifying the best action f and treating the data of the population as given, the firm could aim to *steer* the population towards data that it prefers. To understand the role of steering we focus on prediction. Here, the goal of the firm is to choose a predictive model f that suffers small loss $\ell(f(x), y)$ measured over instances (x, y). The *ex-post risk* that the firm suffers after deploying f on a population $\mathcal{U}$ is given by

$$\frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \ell(f(x_f(u)), y_f(u)) \,. \tag{1}$$

The phenomenon that a predictor can impact the underlying data distribution it aims to predict has been formalized in the work on *performative prediction* by Perdomo et al. [2020]. Expression (1) constitutes an instance of what they call the *performative risk* of a predictor. A key conceptual device in performative prediction is the distribution map $\mathcal{D}(\theta)$ that maps a predictive model, characterized by model parameters θ, to a distribution over data instances. To express our setting within this framework, we assume the firm's action space is to choose a parameter configuration θ for its predictor f_θ from an action space $\mathcal{F} = \Theta$. Let a data instance correspond to $z(u) = (x(u), y(u))$ for $u \in \mathcal{U}$ so we can capture performativity in the features as well as in the labels. Then, the *aggregate distribution* over data $\mathcal{D}(\theta)$ corresponds to the distribution over the potential outcome variable $z_\theta(u)$ after the firm takes action $\theta \in \Theta$, where the randomness comes from u being uniformly drawn from $\mathcal{U}$ as well as randomness in the potential outcomes. The firm's ex-post risk (1) from deploying predictor f_θ corresponds to the performative risk:

$$\mathrm{PR}(\theta) := \mathop{\mathbb{E}}_{z \sim \mathcal{D}(\theta)}[\ell(\theta;\, z)]$$

In prediction problems, the loss typically corresponds to the mismatch between the predicted label and the true label: $\ell(\theta;\, z) = \ell(f_\theta(x), y)$ for $z = (x, y)$.

Ex-ante vs. ex-post optimization. In performative risk minimization, observe that θ arises in two places in the objective: in the distribution $\mathcal{D}(\theta)$ and in the loss $\ell(\theta;\, z)$. To disentangle these effects, we use the notion of the *decoupled risk* [Perdomo et al., 2020]

$$\mathrm{R}(\phi, \theta) := \mathop{\mathbb{E}}_{z \sim \mathcal{D}(\phi)} \ell(\theta;\, z) \,. \tag{2}$$

The decoupled risk denotes the loss of a model θ on the distribution $\mathcal{D}(\phi)$. In this notation, the performative risk $\mathrm{PR}(\theta)$ is equal to $\mathrm{R}(\theta, \theta)$. Moreover, for any choice of model ϕ, we can decompose the performative risk $\mathrm{PR}(\theta)$ as:

$$\mathrm{PR}(\theta) = \mathrm{R}(\phi, \theta) + (\mathrm{R}(\theta, \theta) - \mathrm{R}(\phi, \theta)) \tag{3}$$

This seemingly trivial decomposition reveals the salient difference between ex-ante and ex-post optimization.

Ex-ante optimization focuses on predicting from historical data. Let ϕ by any previously chosen model, then employing supervised learning on historical data sampled from $\mathcal{D}(\phi)$ corresponds to minimizing the first term in the decomposition (3). For any ϕ, the resulting minimizer can be computed statistically and corresponds to $\theta_{\mathrm{SL}} = \arg\min_{\theta \in \Theta} \mathrm{R}(\phi, \theta)$. In the context of digital content recommendation, this ex-ante approach corresponds to training a predictor on the participant data collected on the platform at any previous point in time.

In contrast, *ex-post optimization* accounts for the impact of the model on the distribution, trades-off the two terms in (3), and directly optimizes the performative risk according to $\theta_{\mathrm{PO}} = \arg\min_{\theta \in \Theta} \mathrm{PR}(\theta)$. Solving this problem exactly, and finding the performative optimum θ_{PO} requires optimization over the distribution map $\mathcal{D}(\theta)$. This can either be achieved by explicitly modeling the distribution shifts [Hardt et al., 2016] or by targeted exploration [e.g., Jagadeesan et al., 2022]. Moreover, ex-post optimization is implicitly pursued by any firm that applies A/B testing to evaluate models based on ex-post performance.

It holds that $\mathrm{PR}(\theta_{\mathrm{PO}}) \leq \mathrm{PR}(\theta_{\mathrm{SL}})$, because in ex-post optimization the firm can benefit from intentionally steering the population towards a distribution that it prefers. In the context of prediction this would be a distribution that is easier to predict—for example, a distribution with better Bayes optimal risk. Ex-post predictability as a strong incentive for a firm monetizing "prediction products" has been discussed in various works, including [Russell, 2019].

3.1 Gain of ex-post optimization is bounded by a firm's performative power

We show that the gain of ex-post optimization over ex-ante optimization can be bounded by the firm's performative power with respect to the set of actions Θ and the data vector $z = (x, y)$. Intuitively, if the firm's performative power is low, then the distributions $\mathcal{D}(\theta)$ and $\mathcal{D}(\phi)$ for any $\theta, \phi \in \Theta$ are close to one another. This distributional closeness, coupled with a regularity assumption on the loss, means that the second term in (3) should be small. Thus, using the ex-ante approach of minimizing the first term produces a near-optimal ex-post solution, as we demonstrate in the following result:

Theorem 1. *Let* P *be the performative power of a firm with respect to the action set* Θ*. Let* L_z *be the Lipschitzness of the loss in* z *with respect to the metric* dist*. Let* θ_{PO} *be the ex-post solution and* θ_{SL} *be the ex-ante solution computed from* $\mathcal{D}(\phi)$ *for any past deployment* $\phi \in \Theta$*. Then, we have that:*

$$\mathrm{PR}(\theta_{\mathrm{SL}}) \leq \mathrm{PR}(\theta_{\mathrm{PO}}) + 4L_z\mathrm{P}.$$

If ℓ *is* γ*-strongly convex, the distance in parameter space is at most* $\|\theta_{\mathrm{SL}} - \theta_{\mathrm{PO}}\|_2 \leq \sqrt{\frac{8L_z\mathrm{P}}{\gamma}}$.

Theorem 1 illustrates that a firm with small performative power cannot do much better than ex-ante optimization. Thus, such a firm might be better off sticking to classical supervised learning practices instead of engaging with ex-post optimization that comes with high modeling or experimentation costs. Returning to our example of digital content platforms, this means that if a firm has low performative power, then it loses very little by solving the static problem of predicting target variables from historical data, rather than explicitly trying to shape the videos produced by content creators.

3.2 Ex-post optimization in an economy of predictors

The result in Theorem 1 studies the optimization strategy of a single firm in isolation. In this section, we investigate the interaction between the strategies of multiple firms that optimize simultaneously over the same population. We consider an idealized marketplace where C firms all engage in ex-post optimization and we assume all exogenous factors remain constant. Let $\mathcal{D}(\theta^1, \ldots, \theta^{i-1}, \theta^i, \theta^{i+1}, \ldots, \theta^C)$ be the distribution over $z(u)$ induced by each firm $i \in [C]$ selecting f_{θ^i}. We say a set of predictors $[f_{\theta^1}, \ldots, f_{\theta^C}]$ is a *Nash equilibrium* if and only if no firm has an incentive to unilaterally deviate from their predictor using ex-post optimization:

$$\theta^i \in \arg\min_{\theta \in \Theta} \mathop{\mathbb{E}}_{z \sim \mathcal{D}(\theta^1, \ldots, \theta^{i-1}, \theta, \theta^{i+1}, \ldots, \theta^C)} [\ell_i(\theta;\, z)]$$

Here ℓ_i denotes the loss function chosen by firm i. We show at the Nash equilibrium, the suboptimality of each predictor f_{θ^i} on the induced distribution depends on the firm's performative power.

Proposition 1. *Suppose that the economy is in a Nash equilibrium* $(\theta^1, \ldots, \theta^C)$*, and firm* i *has performative power* P_i *with respect to the action set* Θ*. Let* L_z *be the Lipschitzness of the loss* ℓ_i *in* z *with respect to the metric* dist*. Then, it holds that:*

$$\mathop{\mathbb{E}}_{z \sim \mathcal{D}} [\ell_i(\theta^i;\, z)] \leq \min_{\theta} \mathop{\mathbb{E}}_{z \sim \mathcal{D}} [\ell_i(\theta;\, z)] + L_z \mathrm{P}_i\,,$$

where $\mathcal{D} = \mathcal{D}(\theta^1, \ldots, \theta^C)$ *is the distribution induced at the equilibrium. If* ℓ_i *is* γ*-strongly convex, then we can also bound the distance between* θ^i *and* $\arg\min_\theta \mathbb{E}_{z \sim \mathcal{D}}[\ell_i(\theta;\, z)]$ *in parameter space.*

Proposition 1 implies that if the performative power of the firm is small ($\mathrm{P}_i \to 0$), then the equilibrium becomes indistinguishable from that of a *static*, non-performative economy with distribution $\mathcal{D}$ over content. However, if the firms were to collude—for example, because of common ownership[3]—then they would be able to significantly shift the distribution. In particular, even if the performative power P_i of any given firm i is small, the aggregate performative power of a set of firms S can be much larger, potentially exceeding the aggregate power of the firms constituting S. As a result, low performative power only makes the optimization problem static if firms do not collude.

To illustrate this, we consider a *mixture economy*, where all of the firms share a common loss function ℓ and performative power is uniformly distributed across firms. We defer the formal definition of a mixture economy to Appendix E.4. Informally speaking, let $\mathcal{D}^M(\theta)$ be the distribution map associated with the counterfactual outcomes after deploying of θ in a hypothetical monopoly economy where a single firm holds all the performative power. We can apply Proposition 1 to relate the equilibria in the limit as $C \to \infty$ to properties of the monopoly distribution map $\mathcal{D}^M(\theta)$.

[3] Common ownership refers to the situation where many competitors are jointly held by a small set of large institutional investors [Azar et al., 2018].

Corollary 1. *Suppose that all firms share the same loss function $\ell_i = \ell$ for all i. Let θ^* be a symmetric equilibrium in the mixture economy with C platforms. As $C \to \infty$, it holds that:*

$$\mathop{\mathbb{E}}_{z\sim\mathcal{D}^M(\theta^*)}[\ell(\theta^*;\, z)] = \mathop{\mathbb{E}}_{z\sim\mathcal{D}(\theta^*,\ldots,\theta^*)}[\ell(\theta^*;\, z)] \to \min_\theta \mathop{\mathbb{E}}_{z\sim\mathcal{D}(\theta^*,\theta^*,\ldots,\theta^*)}[\ell(\theta;\, z)] = \min_\theta \mathop{\mathbb{E}}_{z\sim\mathcal{D}^M(\theta^*)}[\ell(\theta;\, z)].$$

Corollary 1 demonstrates that a symmetric equilibrium approaches a *performatively stable point* of $\mathcal{D}^M$. In contrast, if a set of firms S collude, then the performative effect that they induce by all taking the same action is $|S|$ times as large as the performative effect that any single firm could have induced with that action. Thus, their aggregate performative power is $\sum_{i\in S} P_i$. If all of the firms collude, they would benefit from all choosing a *performatively optimal point* of $\mathcal{D}^M$—which is also the equilibrium in a monopoly economy with a single firm. Since performatively optimal and performatively stable points can be far apart in general [Miller et al., 2021], a competitive economy can exhibit a significantly different equilibrium from that of the monopoly or collusive economy.

4 Performative power in concrete settings

Having elucidated the role of performative power in the firm's optimization strategy, we now turn to a stylized market model and investigate how performative power depends on the economy in which the firm operates. We use *strategic classification* [Hardt et al., 2016] as a test case for our definition. In strategic classification, participants strategically adapt their features with the goal of achieving a favorable classification outcome. Hence, performative power is determined by the degree to which a firm's classifier can impact participant features. Our mode of investigation is to use this concrete market setting to examine whether performative power exhibits the qualitative behavior that we would expect from a measure of market power in the presence of competition and outside options.

4.1 Strategic classification setup

Let $x(u)$ be the *features* and $y(u)$ the *binary label* describing a participant $u \sim \mathcal{U}$. A firm chooses a binary predictor $f : \mathbb{R}^m \to \{0, 1\}$ and incurs loss $\ell(f(x), y) = |f(x) - y|$. The shifts in the population of participants in response to a change in the firm's predictor f is assumed to arise from participants strategically adjusting their features x in an attempt to achieve a better prediction outcome $f(x)$. Let $\mathcal{D}_{\text{base}}$ denote the base distribution over features and labels $(x_{\text{orig}}(u), y_{\text{orig}}(u))$ absent any strategic adaptation, which we assume is continuous and supported everywhere. Let $\mathcal{D}(f)$ be the distribution over potential outcomes $(x_f(u), y_f(u))$ that arises from deploying model f. We assume that participant u incurs a cost $c(x_{\text{orig}}(u), x)$ for changing their features to x. We further assume that c is a metric and that any feature change that deviates from the original features results in nonnegative cost for participants, but does not change the label $y_f(u)$ from the original label $y_{\text{orig}}(u)$. The model is non-stateful and the cost for feature changes is measured relative to the *original* features.

Since performative effects surface as changes in participant features in strategic classification, we measure performative power over $z(u) = x(u)$. The choice of distance metric over $\mathcal{X} \times \mathcal{X}$ enables us to define how to weight specific feature changes. In the digital content recommendation example, if we are interested in the burden on *content creators*, an ideal distance metric would be aligned with the cost function c. However, if we are interested in measuring the impact of changes in content on viewers, a distance metric that more closely reflects the harm of specific content change for viewers might be more appropriate. We will keep this distance metric abstract in our analysis. (In Appendix C.1, we discuss more details of how performative power might be instantiated for digital content recommendation and hiring platforms).

4.2 Performative power in the monopoly setting

We start by considering the typical strategic classification setup proposed by Hardt et al. [2016], where there is a single firm that offers utility γ to its participants for a positive classification. An implicit assumption in this classical setting is that participants want to use the service independent of the classifier being deployed and the surplus utility corresponds to $\Delta\gamma = \gamma$. To better reflect real market situations, we incorporate *outside options*: alternative services that are available to participants without any extra effort and offer utility level $\beta \geq 0$. Outside options decrease the budget participants are willing to invest to $\Delta\gamma = \max(0, \gamma - \beta)$.

We adopt the following standard rationality assumption on participant behavior.

Assumption 1 (Participant Behavior Specification)**.** *Let $\Delta\gamma \geq 0$ be the surplus utility that a participant can expect from a positive classification outcome from classifier f over any outside option. Then, a participant $u \in \mathcal{U}$ with original features $x_{orig}(u)$ will change their features according to*

$$\arg\max_{x'} \left(\Delta\gamma f(x') - c(x_{orig}(u), x')\right).$$

Assumption 1 guarantees that a participant will change their features if and only if the cost of a feature change is no larger than $\Delta\gamma$. Furthermore, if a participant change their features, then they will expend the minimal cost required to achieve a positive outcome. This specification of participant behavior restricts the set of potential values that $x_f(u)$ to be close to the participant's original features $x_{\text{orig}}(u)$. We let $\mathcal{X}$ be the set of potential values of $x_f(u)$, given by:

$$\mathcal{X}(u; \Delta\gamma) := \{x : c(x_{\text{orig}}(u), x) \leq \Delta\gamma\}.$$

Given this constraint set on the counterfactual data points, we can bound performative power:

Lemma 1. *The performative power* P *of the firm with respect to any set of predictors $\mathcal{F}$ satisfies* $\mathrm{P} \leq \sup_{u\in\mathcal{U}} \operatorname{diam}(\mathcal{X}(u; \Delta\gamma))$*, where the diameter is measured with respect to the metric* dist.

The set $\mathcal{X}$ is defined with respect to the cost function c whereas the diameter of the set is measured in terms of the metric dist.[4] To relate the two measures we define

$$L := \sup_{x,x'} \frac{\mathrm{dist}(x, x')}{c(x, x')}.$$

be the Lipschitz constant of dist with respect to c. This constant will show up in the next bound and can be evaluated for any choice of distance metric.

Corollary 2. *The performative power of a firm in the monopoly setup can be bounded as* $\mathrm{P} \leq 2L\,\Delta\gamma$*, where $\Delta\gamma$ measures the surplus utility offered by the service of the firm over outside options.*

This result demonstrates that $\Delta\gamma > 0$ is a prerequisite for a firm to have any performative power, even in a monopoly economy. This qualitative behavior of performative power is in line with common intuition in economics that monopoly power relies on the firm offering a service that is superior to existing options. Interestingly, in the literature, it is typically assumed that $\Delta\gamma = \gamma$ and the firm is able to extract up to the full utility from participants close to the decision boundary. This means, in the worst case, participants invest effort but end up with a net utility of 0 in the classical setup. Outside options, however, restrict the firm's ability to extract this level of utility from participants, thereby reducing their performative power. While we have focused on the impact of participants in the worst-case thus far, in Appendix C.2, we reason about the *heterogeneity of the participant base* and its performative power in the monopoly setting.

4.3 Firms competing for participants

We next consider a dynamic model of *competition* where participants choose the firm that offers higher utility. In this model of perfectly elastic demand, we demonstrate how the presence of competition significantly reduces the performative power of a firm. We will show that in an economy with two firms, each firm's performative power can drop to zero at equilibrium. This qualitative behavior of performative power is in line with well-known results on market power under Bertrand competition in economics, see e.g., [Baye and Kovenock, 2008].

To model competition in strategic classification, we specify participant behavior as follows. As before, for a given predictive model f, participant u follows standard microfoundations (following Assumption 1). It remains to specify how each participant *selects* a firm. We assume that they select the firm for which they receive a higher utility. That is, if the first firm deploys f_1 and the second firm deploys f_2, then they will choose f_1 if

$$\max_{x'} \left(\gamma f_1(x') - c(x_{\text{orig}}(u), x')\right) > \max_{x'} \left(\gamma f_2(x') - c(x_{\text{orig}}(u), x')\right),$$

and choose f_2 otherwise. We assume that a participant tie-breaks in favor of the lower threshold, randomizing if the two thresholds are equal. After choosing firm $i \in \{1, 2\}$, they change their features according to $x_f(u) = \arg\max_{x'} \left(\gamma f_i(x') - c(x_{\text{orig}}(u), x')\right)$.

[4] We keep the distance metric dist abstract in our analysis. See Appendix C.1 for a discussion.

For the results in this section, we focus on a 1-dimensional setup where $\mathcal{F}$ is the set of threshold functions. We assume that the cost function $c(x, x')$ is continuous in both of its arguments, strictly increasing in x' for $x' > x$, strictly decreasing for $x' < x$, and satisfies $\lim_{x' \to \infty} c(x, x') = \infty$. Furthermore, we assume that the posterior $p(x) = \mathbb{P}_{\mathcal{D}_{\text{base}}}[Y = 1 \mid X = x]$ satisfies the following conditions: it is strictly increasing in x with $\lim_{x \to -\infty} p(x) = 0$, and $\lim_{x \to \infty} p(x) = 1$.

Proposition 2. *Suppose that the economy is at a symmetric state where both firms choose the same classifier θ. Let θ_{min} be the minimum threshold classifier in $\mathcal{F}(\theta)$, and for $\theta' \in \{\theta_{min}, \theta\}$, let θ'_M be the unique value such that $c(\theta'_M, \theta') = \gamma$ and $\theta'_M < \theta'$. Then, the performative power with respect to an action set $\mathcal{F}(\theta) \subseteq \mathcal{F}$ is upper bounded by:*

$$P \leq L \min(c(\theta_{min}, \theta), \gamma) + L\gamma \cdot \mathbb{P}_{\mathcal{D}_{base}}[x \in [(\theta_{min})_M, \theta_M)].$$

The performative power of a firm thus purely arises from how much larger the market threshold is than the minimum threshold a firm can deploy within their action set. We introduce the constraint $\mathcal{F}(\theta)$ on the action set of the firm in order to capture that not all actions are realistic for the firm to select. For example, one way that a firm could gain all of the participants and induce significant performative effects is by setting its threshold so that it accepts all participants. However, in doing so, the firm would accept many unqualified participants, and their profit or utility could become negative. Thus, to get a practically relevant measure of performative power, we constrain the set of actions $\mathcal{F}(\theta)$ to contain only predictors that result in a *nonnegative profit*, given that θ is chosen by the other firm. When working with historical data to evaluate performative power the choice of classifiers, we can observe naturally fall within this set, given the economic incentives of a firm.

Competition for positively labeled participants. Under a natural specification of the firm utility functions, the performative power is driven down to zero in the presence of a single competitor. The intuition is that at the Nash equilibrium of the economy, the market threshold is exactly the minimum threshold that the firm can deploy within their action set. It thus follows from Proposition 2 that the performative power of the firm is 0. We defer the details of this derivation to Appendix C.3.

5 Discussion

Performative power offers a tool to assess the power of a firm within a digital economy that does not require precisely specifying market dynamics or participant behavior. This approach to measuring power is conceptually different from classical notions of market power in economics that rely on a very specific market definition and have shortcomings in the context of digital economies (see Appendix D for a discussion). Furthermore, performative power is directly linked to the ability of a firm to employ steering to reduce the risk of its predictor, a strategy that bears potential for user harm. A similar moral concern has been raised by Shmueli and Tafti [2020] based on the fact that firms engage in behavioral modification for the sake of improving their prediction products. For this reason, we believe that performative power constitutes a relevant measure of power in the context of antitrust enforcement in digital economies.

To connect performative power to anti-trust enforcement, the parameters of the definition should be chosen to capture harm to the impacted population of participants. We envision that the attributes $z(u)$ should consist of sensitive features that are maximally impacted by the firm, and the distance function should be roughly aligned with the utility function of participants. Ultimately, antitrust enforcement is concerned with manifested harm to the participants. High performative power signals the potential for harm and exploitation, but does not necessarily imply harm. For example, in the special case where firm's objective happens to perfectly align with participant utility, steering could lead to improvement. As a result, performative power on its own is not a measure of participant harm, but rather an indicator of *potential* harm that merits investigation.[5]

More broadly, with this work we hope to provide a formal basis to encourage an emerging discourse on the important topic of power in predictive systems within the machine learning community.

[5] The report of the European Commission on competition policy for the digital era [Crémer et al., 2019] suggests putting the burden of proof on the firm to provide a solid efficiency defence.

References

Jose Azar, Martin C. Schmalz, and Isabel Tecu. Anticompetitive effects of common ownership. *The Journal of Finance*, 73(4):1513–1565, 2018.

Ian Ball. Scoring strategic agents. *ArXiv:1909.01888*, 2020.

Michael R. Baye and Dan Kovenock. *Bertrand competition*. Palgrave Macmillan UK, London, 2008.

Yahav Bechavod, Chara Podimata, Zhiwei Steven Wu, and Juba Ziani. Information discrepancy in strategic learning. *Arxiv:2103.01028*, 2021.

Jean Magnan de Bornier. The "Cournot-Bertrand Debate": A Historical Perspective. *History of Political Economy*, 24(3):623–656, 09 1992.

Mark Braverman and Sumegha Garg. The role of randomness and noise in strategic classification. In *Proc. 1st FORC 2020*, volume 156, pages 9:1–9:20, 2020.

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

Michael Brückner, Christian Kanzow, and Tobias Scheffer. Static prediction games for adversarial learning problems. *JMLR*, 13(1):2617–2654, September 2012.

Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, pages 191–198, 2016.

Jacques Crémer, Yves-Alexandre de Montjoye, and Heike Schweitzer. *Competition Policy for the digital era : Final report*. Publications Office of the European Union, 2019.

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, 2018.

Roy Dong and Lillian J Ratliff. Approximate regions of attraction in learning with decision-dependent distributions. *arXiv preprint arXiv:2107.00055*, 2021.

B. J. Fogg. Persuasive technology: Using computers to change what we think and do. *Ubiquity*, 2002.

Alex Frankel and Navin Kartik. Muddled Information. *Journal of Political Economy*, 127(4): 1739–1776, 2019.

Alex Frankel and Navin Kartik. Improving Information via Manipulable Data. *Working Paper*, 2020.

Ganesh Ghalme, Vineet Nair, Itay Eilat, Inbal Talgam-Cohen, and Nir Rosenfeld. Strategic classification in the dark. In *Proceedings of the 38th International Conference on Machine Learning*, pages 3672–3681. PMLR, 2021.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. ITCS '16, pages 111–122, 2016.

Christopher Hennessy and Charles Goodhart. Goodhart's law and machine learning. *SSRN*, 2020.

Lily Hu, Nicole Immorlica, and Jennifer Wortman Vaughan. The disparate effects of strategic manipulation. In *Proc. FAccT*, pages 259–268, 2019.

Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: Performative gradient descent. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 4641–4650, 2021.

Meena Jagadeesan, Celestine Mendler-Dünner, and Moritz Hardt. Alternative microfoundations for strategic classification. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 4687–4697. PMLR, 2021.

Meena Jagadeesan, Tijana Zrnic, and Celestine Mendler-Dünner. Regret minimization with performative feedback. *Arxiv:2202.00628*, 2022.

Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, pages 825–844, 2019.

A. P. Lerner. The concept of monopoly and the measurement of monopoly power. *The Review of Economic Studies*, 1(3):157–175, 1934.

Sagi Levanon and Nir Rosenfeld. Strategic classification made practical. *Arxiv:2103.01826*, 2021.

Qiang Li and Hoi-To Wai. State dependent performative prediction with stochastic approximation. *ArXiv:2110.00800*, 2021.

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.

John Miller, Juan C. Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. *Arxiv:2102.08570*, 2021.

Smitha Milli, John Miller, Anca D. Dragan, and Moritz Hardt. The social cost of strategic classification. In *Proc. FAccT*, pages 230–239, 2019.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian J. Ratliff. Multiplayer performative prediction: Learning in decision-dependent games. *ArXiv:2201.03398*, 2022.

Sridhar Narayanan and Kirthi Kalyanam. Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science*, 34(3):388–407, may 2015. ISSN 1526-548X.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proc. 37th ICML*, volume 119, pages 7599–7609, 2020.

Georgios Piliouras and Fang-Yi Yu. Multi-agent performative prediction: From global stability and optimality to chaos. *Arxiv:2201.10483*, 2022.

Mitas Ray, Lillian J Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. 2022.

S. Russell. *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin Publishing Group, 2019.

Galit Shmueli and Ali Tafti. "Improving" prediction of human behavior using behavior modification. *Arxiv:2008.12138*, 2020.

Stigler Committee. Final report: Stigler committee on digital platforms. available at https://research.chicagobooth.edu/stigler/media/news/committee-on-digitalplatforms-final-report, September 2019.

Chad Syverson. Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, 33(3):23–43, August 2019.

Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, 2008.

Killian Wood, Gianluca Bianchin, and Emiliano Dall'Anese. Online projected gradient descent for stochastic optimization with decision-dependent distributions. *IEEE Control Systems Letters*, 6: 1646–1651, 2022.

Tijana Zrnic, Eric Mazumdar, Shankar Sastry, and Michael Jordan. Who leads and who follows in strategic classification? *Advances in Neural Information Processing Systems*, 34, 2021.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]