Making Decisions Under Outcome Performativity

Michael P. Kim
Miller Institute, UC Berkeley, CA, USA

Juan C. Perdomo
Department of Computer Science, UC Berkeley, CA, USA

Abstract
Decision-makers often act in response to data-driven predictions, with the goal of achieving favorable outcomes. In such settings, predictions don’t passively forecast the future; instead, predictions actively shape the distribution of outcomes they are meant to predict. This performative prediction setting [28] raises new challenges for learning “optimal” decision rules. In particular, existing solution concepts do not address the apparent tension between the goals of forecasting outcomes accurately and steering individuals to achieve desirable outcomes.

To contend with this concern, we introduce a new optimality concept – performative omniprediction – adapted from the supervised (non-performative) learning setting [9]. A performative omnipredictor is a single predictor that simultaneously encodes the optimal decision rule with respect to many possibly-competing objectives. Our main result demonstrates that efficient performative omnipredictors exist, under a natural restriction of performative prediction, which we call outcome performativity. On a technical level, our results follow by carefully generalizing the notion of outcome indistinguishability [5] to the outcome performative setting. From an appropriate notion of Performative OI, we recover many consequences known to hold in the supervised setting, such as omniprediction and universal adaptability [19].

2012 ACM Subject Classification Theory of computation → Theory and algorithms for application domains

Keywords and phrases performative prediction, outcome indistinguishability

Digital Object Identifier 10.4230/LIPIcs.ITCS.2023.79


Funding Michael P. Kim: MPK is supported by the Miller Institute for Basic Research in Science and, in part, by the Simons Collaboration on Algorithmic Fairness.

Acknowledgements The authors thank Parikshit Gopalan, Moritz Hardt, Celestine Mendler-Dunner, Omer Reingold, and Tijana Zrnic for helpful discussions throughout the development of the project.

1 Introduction

Data-driven predictions inform policy decisions that directly impact individuals. Proponents argue that by understanding patterns from the past, decisions can be optimized to improve future outcomes, to the benefit of individuals and institutions [22]. In the US educational system, for instance, early warning systems (EWS) have become a key tool used by states to combat low graduation rates [1, 30]. The rationale for using such systems is clear. Given a predictor that, for each student, estimates the likelihood of graduation, school districts can identify high-risk students at a young age, directing resources to improve individuals’ outcomes, and in turn, the districts’ graduation rates. Despite compelling arguments, reliably predicting life outcomes remains a largely-unsolved problem in machine learning.

A key challenge in utilizing predictions to inform decisions is that, often, predictions influence the outcomes they’re meant to forecast. In the education example above, districts consider predictions of graduation with the intention of effecting graduation outcomes. In this
Making Decisions Under Outcome Performativity

situation – where predictions determine interventions, which influence outcomes – accuracy can be a paradoxical notion. If a predictor correctly identifies high risk individuals as likely to suffer negative outcomes, after successful interventions, the individuals’ outcomes will be positive and the initial predictions will appear inaccurate. To apply data-driven tools effectively, decision-makers must resolve an apparent tension between the objectives of forecasting individuals’ outcomes reliably and steering individuals to achieve better outcomes.

Recent work of [28] introduced performative prediction to contend with the fact that predictions not only forecast, but also shape the world. Informally, a prediction problem is performative if the act of prediction influences the distribution on individual-outcome pairs. From early warning systems, to online content recommendations, to public health advisories: across many contexts, individuals respond to predictions in a manner that changes the likelihood of possible outcomes (successful graduation, increased click rate, or decreased disease caseload).

In their original work on the subject, [28] frame the goal of performative prediction through loss minimization. In this framing, the ultimate goal is to learn a performatively optimal decision rule. A decision rule \( h_{po} \) is performatively optimal if it achieves the minimal expected loss (within some class of decision rules \( \mathcal{H} \)) over the distribution that it induces,

\[
h_{po} \in \arg \min_{h \in \mathcal{H}} \mathbb{E}_{(x,y) \sim D(h)}[\ell(x, h(x), y)].
\]

Here, \( D(h) \) is the distribution over \((x, y)\) pairs observed as response to deploying \( h \).

For generality’s sake, performative prediction makes minimal restrictions on how the distribution may respond to a chosen decision rule. In particular, the choice to deploy a hypothesis \( h \), may change the joint distribution \((x, y) \sim D(h)\) over individual-outcome pairs, essentially arbitrarily. This generality enables us to write a broad range of prediction problems – including supervised learning [29], strategic classification [11], and causal inference [25] – as special cases of performative prediction. In all, [28] establishes a powerful framework for reasoning about settings where the distribution of examples responds to the predictions.

While powerful, the framework has two noticeable limitations. First, achieving performative optimality is hard. Without any assumptions on the distributional response \( D(\cdot) \), achieving performative optimality requires exhaustive search over the hypothesis class \( \mathcal{H} \). Furthermore, even under strong structural assumptions on the distributional response and choice of loss \( \ell \), it is known that convex optimization does not suffice to achieve optimality [28, 26]. Stated another way: the generality of performative prediction does not come for free. To date, all existing methods for performative optimality require strong specification assumptions on the outcome distribution and distributional response.

The second limitation arises due to formulating performative prediction as a loss minimization problem: the loss \( \ell \) is fixed, once and for all. In performative prediction, different losses can encode drastically different objectives: losses are used not only to promote accuracy of predictions, but also to encourage favorable outcome distributions. Consider a loss designed for accurate forecasting, e.g., the squared error \((\hat{y} - y)^2\). In this case, the optimal decision rule will prioritize accuracy without regard for the “quality” of the outcome distribution. On the other hand, consider a loss designed to steer towards positive outcomes, \(1 - y\). Here, there is no notion of accuracy (the loss ignores the prediction \( \hat{y} \)), but instead, the objective is to nudge the distribution of outcomes towards \( y = 1 \).

\[\text{[28] assume only a Lipschitzness condition, where similar hypotheses } h \text{ and } h' \text{ give rise to similar distributions } D(h) \text{ and } D(h'), \text{ measured in Wasserstein (earth mover’s) distance.}\]
Encoding the decision-making objective through a single loss function forces the learner to choose the “correct” objective at train time. Downstream decision-makers, however, may reasonably want to explore different objectives according to their own sense of “optimality”. In the existing formulations for performative prediction, exploring different losses requires re-training from scratch. In this work, we investigate an alternative formulation that enables decision-makers to efficiently explore optimal decision rules under many different objectives.

1.1 Decision-Making under Outcome Performativity

To begin, we introduce a refinement of the performative prediction setting, which we call outcome performativity. Outcome performativity focuses on the effects of local decisions on individuals’ outcomes, rather than the effect of broader policy on the distribution of individuals. For instance, our example of graduation prediction is modeled well by outcome performativity. For a given a student, the EWS prediction they receive affects their future graduation outcome, but does not influence their demographic features or historical test scores. In other words, we narrow our attention to the performative effects of decisions $h(x)$ on the conditional distribution over outcomes $y$, rather than the effects of the decision rule $h$ on the distribution as a whole $D(h)$. This reframing of performativity still captures many important decision-making problems, but gives us additional structure to address some of the limitations in the original formulation.

On a technical level, outcome performativity imagines a data generating process over triples $(x,\hat{y},y^\ast)$ where $x \sim D$ is sampled from a static distribution over inputs, then a prediction or decision $\hat{y} \in \hat{Y}$ is selected (possibly as a function of $x$), and finally the true outcome $y^\ast \in Y$ is sampled conditioned on $x$ and $\hat{y}$. We focus on binary outcomes $Y = \{0, 1\}$.

In this setting, the outcome performativity assumption posits the existence of an underlying probability function, $p^\ast : X \times \hat{Y} \to [0, 1]$, where for a given individual $x \in X$ and decision $\hat{y} \in \hat{Y}$, the true outcome $y^\ast$ is sampled as a Bernoulli with parameter $p^\ast(x, \hat{y})$. We refer to the true outcome distribution $p^\ast$ as Nature.

By asserting a fixed “ground truth” probability function, the outcome performativity framework does not allow for arbitrary distributional responses and limits the generality of the approach. For instance, outcome performativity does not capture strategic classification. But importantly, by refining the model of performativity, there is hope that we may sidestep the hardness results for learning optimal performative predictors.

1.1.1 Performative Omniprediction

We begin by observing that under outcome performativity, the true probability function $p^\ast$ suggests an optimal decision rule $f^\ast_\ell : X \to \hat{Y}$ for any loss $\ell$. In our setting, $p^\ast$ governs the outcome distribution, so given an input $x \in X$, the optimal decision $f^\ast_\ell(x)$ is determined by a simple, univariate optimization procedure over a discrete set $\hat{Y}$:

$$f^\ast_\ell(x) \in \arg\min_{\hat{y} \in \hat{Y}} E_{y^\ast \sim p^\ast(x, \hat{y})}[\ell(x, \hat{y}, y^\ast)].$$

(2)

In general, outcome performativity could be defined for larger outcome domains. Handling such domains is possible, but technical. We restrict our attention to binary outcomes to focus on the novel conceptual issues.
Note that the decision rule $f^*_\ell(x)$ minimizes the loss pointwise for $x \in \mathcal{X}$. Consequently, averaging over any static, feature distribution $\mathcal{D}$, the decision rule $f^*_\ell$ is performative optimal for any hypothesis class $\mathcal{H}$, loss $\ell$, and marginal distribution $\mathcal{D}$:

$$
\mathbb{E}_{y^* \sim p^*(x,f^*_\ell(x))} [\ell(x,f^*_\ell(x),y^*)] \leq \min_{h \in \mathcal{H}} \mathbb{E}_{y^* \sim p^*(x,h(x))} [\ell(x,h(x),y^*)].
$$

While the existence of $p^*$ implies the existence of optimal decision rules under outcome performativity, we make no assumptions about the learnability of $p^*$. In general, the function $p^*$ may be arbitrarily complex, so learning (or even representing!) $p^*$ may be infeasible, both computationally and statistically.

Still, the above analysis reveals the power of modeling the probability function $p^* : \mathcal{X} \times \hat{\mathcal{Y}} \rightarrow [0, 1]$. The optimal probability function $p^*$ encodes the optimal decision rule $f^*_\ell$ for every loss function $\ell$. This perspective raises a concrete technical question: short of learning $p^*$, can we learn a probability function $\hat{p} : \mathcal{X} \times \hat{\mathcal{Y}} \rightarrow [0, 1]$ that suggests an optimal decision rule, via simple post-processing, for many different objectives? Recent work of [9] studied the analogous question in the context of supervised learning (without performativity), formalizing a solution concept which they call omniprediction. Intuitively, an omnipredictor is a single probability function $\hat{p}$ that suggests an optimal decision rule for many different loss functions $\mathcal{L}$.

In this work, we generalize omniprediction to the outcome performativity setting. As a solution concept, performative omniprediction directly addresses the limiting assumption in performative prediction that the loss $\ell$ is known and fixed. Given a performative omnipredictor, a decision-maker can explore the consequences of optimizing for different losses, balancing the desire for forecasting and steering, as they see fit. Technically, given a predictor $\hat{p}$, we define $\hat{f}_\ell : \mathcal{X} \rightarrow \hat{\mathcal{Y}}$ to be the optimal decision rule, that acts as if outcomes are governed by $\hat{p}$.

$$
\hat{f}_\ell(x) \in \arg\min_{\hat{y} \in \hat{\mathcal{Y}}} \mathbb{E}_{\hat{y} \sim \hat{p}(x,\hat{y})} [\ell(x,\hat{y},\hat{y})]
$$

We emphasize that, for any loss $\ell$, the decision rule $\hat{f}_\ell(x)$ is an efficient post-processing of the predictions given by $\hat{p}(x,\hat{y})$ for $\hat{y} \in \hat{\mathcal{Y}}$. A performative omnipredictor is a model of nature $\hat{p} : \mathcal{X} \times \hat{\mathcal{Y}} \rightarrow [0, 1]$ that induces a corresponding decision rule $\hat{f}_\ell$ that is performatively optimal over a collection of losses $\ell \in \mathcal{L}$.

**Definition 1 (Performative Omnipredictor).** For a collection of loss functions $\mathcal{L}$, hypothesis class $\mathcal{H}$, and $\varepsilon \geq 0$, a predictor $\hat{p} : \mathcal{X} \times \hat{\mathcal{Y}} \rightarrow [0, 1]$ is an $(\mathcal{L}, \mathcal{H}, \varepsilon)$-performative omnipredictor for an input distribution $\mathcal{D}$ if for every $\ell \in \mathcal{L}$, the decision rule $\hat{f}_\ell$ is $\varepsilon$-performative optimal over $\mathcal{H}$.

$$
\mathbb{E}_{y^* \sim p^*(x,f_\ell(x))} [\ell(x,\hat{f}_\ell(x)),y^*)] \leq \arg\min_{h \in \mathcal{H}} \mathbb{E}_{y^* \sim p^*(x,h(x))} [\ell(x,h(x),y^*)] + \varepsilon
$$

(3)
While an intriguing prospect, omniprediction is particularly ambitious in the performative world. Whereas most supervised learning losses have roughly the same goal (to accurately forecast the outcome), losses in the performative world can encode entirely contradictory objectives. For instance, we can define a pair of losses $\ell_0$ and $\ell_1$ that reward decisions that steer outcomes to be 0 and 1, respectively. A performative omnipredictor must contend with these contradictions, providing optimal decision rules under performative effects.

Concretely, under outcome performativity, there is a certain circularity in naively determining the optimal decision $\tilde{f}(x)$ from a prediction $\tilde{p}(x, \hat{y})$. Choosing an “optimal” decision $\tilde{f}(x)$ causes a shift in the distribution on the outcome $y^* \sim p^*(x, \tilde{f}(x))$, which may imply a different “optimal” decision, which seems to lead to a continuing cycle of dependency. In this way, any performative omnipredictor $\tilde{p}$ must encode the optimal decision rule $\tilde{f}_\ell$ for each $\ell \in \mathcal{L}$, anticipating the shift induced by the choice of $\tilde{f}_\ell$. In this work, we ask whether – despite this key challenge – efficient performative omnipredictors exist, and if so, can we learn them?

1.2 Our Contributions

Our first contributions are conceptual, introducing the outcome performativity setting and the notion of performative omnipredictors. As an abstraction, outcome performativity strikes a balance with enough generality to model many real-world phenomena paired with enough structure to give effective solutions. For settings where the distributional response occurs predominantly as outcome performativity, the framework is well-scoped to contend with the challenges of performative prediction. In particular, performative omnipredictors provide an effective solution concept to address the tension between different objectives under performativity. With these conceptual contributions in place, we turn to the feasibility of omniprediction under outcome performativity.

1.2.1 Efficient Performative Omnipredictors Exist

Our first technical contribution demonstrates existence of efficient performative omnipredictors. We prove that for any class of losses $\mathcal{L}$ and any hypothesis class $\mathcal{H}$, there exists a performative omnipredictor $\tilde{p}$ of complexity that scales polynomially with the complexity of computing the losses and hypotheses.

> Theorem 2. Suppose $\mathcal{D}$ is a fixed distribution over $\mathcal{X}$. Let $\mathcal{L} \subseteq \{\ell : \mathcal{X} \times \hat{\mathcal{Y}} \times \mathcal{Y} \to [0, 1]\}$ be a set of bounded loss functions, and let $\mathcal{H} \subseteq \{h : \mathcal{X} \to \hat{\mathcal{Y}}\}$ be a hypothesis class of decision rules. If the functions in $\mathcal{L}$ and $\mathcal{H}$ can be computed by circuits of size $s$, then there exists a $(\mathcal{L}, \mathcal{H}, \varepsilon)$-performative omnipredictor of circuit complexity $\text{poly}(s, |\hat{\mathcal{Y}}|)/\varepsilon^2$.

Importantly, this result holds for any class of bounded losses. The collection $\mathcal{L}$ may include losses for forecasting and steering, or may include losses that steer towards different outcomes. Still, the predictor $\tilde{p}$ will encode a performative optimal decision rule for each such loss $\ell \in \mathcal{L}$. Furthermore, the complexity of this predictor scales gracefully with the complexity of the losses and hypotheses and the available decisions, independent of the complexity of Nature $p^*$. Even if the true probability function $p^*$ is intractably-complex, there exists a simple function $\tilde{p}$ that mimics the omniprediction behavior, provided the losses and hypotheses are sufficiently simple.
1.2.2 Learning Performative Omnipredictors Reduces to Supervised Learning

In fact, the proof of existence is constructive. We establish the feasibility of performative omnipredictions by devising a boosting-style learning algorithm, inspired by the original algorithm for learning (non-performative) omnipredictors \cite{12, 5, 9}. As in the supervised case, we show that learning omnipredictors reduces to an auditing task. Despite the fact that in performative prediction, different decision rules induce different distributions, we show that – given appropriately randomized data – this auditing task can be solved using only supervised learning primitives implementable in finite samples. That is, under outcome performativity, there is a surprising reduction from the task of learning optimal performative predictors to the task of non-performative supervised learning.

Formally, we assume that the learner has access to a collection of data triples \((x, \hat{y}, y) \sim D_{rc} \) where inputs are sampled from the data distribution \(x \sim D\), decisions \(\hat{y}\) are assigned uniformly at random, and the outcome \(y^* \sim p^*(x, \hat{y})\) is sampled from Nature, for the given individual and randomly-assigned decision. Given an efficiently bounded number of samples access from this distribution, we show how to learn performative omnipredictors assuming access to a supervised learner for the hypothesis class \(H\). We formalize this learning assumption in terms of cost-sensitive classification \cite{7}.

\textbf{Theorem 3 (Informal).} Assume sample access to \(D_{rc}\) and suppose that \(A\) is a cost-sensitive learning algorithm for the hypothesis class \(H\). There is a polynomial-time algorithm, that, for any set of bounded losses \(L\), returns a \((L, H, \varepsilon)\)-performative omnipredictor using at most \(\text{poly}(1/\varepsilon, |\hat{Y}|, \log|H|, \log|L|)\) many samples from \(D_{rc}\) while also making \(|L| \cdot \text{poly}(1/\varepsilon, |\hat{Y}|)\) oracle calls to \(A\).

The guarantees of this algorithm represent a significant point of departure from previous work on performative prediction. Specifically, previous algorithms for learning performatively optimal models (for a single loss) hinged on the condition that predictions had very mild, and highly-structured (e.g. linear) impact on the induced data distributions as in \cite{26, 15}. Conversely, within the outcome performativity restriction, we make no assumptions on the way predictions influence outcomes. Further, the omnipredictor output in the guarantee of Theorem 3 has complexity scaling as stated in Theorem 2. In other words, our learning algorithm makes no “realizability” assumptions and outputs an efficient predictor, regardless of the complexity of Nature.

1.2.3 Universally-Adaptable Omnipredictors

Outcome performativity focuses attention on performative shifts in the outcome distribution as a function of the chosen decision \(\hat{y} \in \hat{Y}\). In particular, it excludes performative effects in the distribution over individuals \(X\). Despite this limitation, our final result shows that we can learn performative omnipredictors that are robust to exogenous (non-performative) shifts in the distribution over individuals.

Adapting the notion of universal adaptability, introduced in the context of statistical estimation by \cite{19}, we show how to learn \textit{universally-adaptable} performative omnipredictors. Whereas performative omnipredictors guarantee optimality on a fixed marginal distribution \(D\) over individuals, universally-adaptable omnipredictors give the same optimality guarantee, simultaneously, over a rich class of input distribution shifts \(D_W\). Each distribution in \(D_\omega \in D_W\) corresponds to the reweighting of probabilities in \(D\) by some importance weight function \(\omega\) in some pre-specified class \(W\).
Theorem 4 (Informal). Let \( L \) be a set of bounded loss functions, \( \mathcal{H} \) a hypothesis class of decision rules, and let \( \mathcal{W} \subseteq \{ \omega : \mathcal{X} \rightarrow [0, \omega_{\text{max}}] \} \). If the functions in \( L \), \( \mathcal{H} \), and \( \mathcal{W} \) can be computed by circuits of size \( s \), then there exists a predictor \( \tilde{p} \), computable by a circuit of size at most \( \text{poly}(s, |\hat{Y}|) \cdot \omega_{\text{max}}^2/\varepsilon^2 \), that is a \((L, \mathcal{H}, \varepsilon)\)-performative omnipredictor for every distribution over individuals \( D_\omega \in D_W \).

The result follows by augmenting the class of loss functions \( L_W \) to account for shifts under \( \mathcal{W} \), and again, applying the constructive learning algorithm from Theorem 3. We emphasize that the learner only needs to account for the class of shifts at training time. At evaluation time, the decision-maker need not know anything about the underlying distribution over individuals. Indeed, the decision-maker can simply use \( \tilde{p} \) as before, post-processing to decisions \( \tilde{f}_\ell(x) \) for any \( \ell \in L \) on an input-by-input basis.

1.3 Our Techniques: Performative Outcome Indistinguishability

We begin our technical overview with a simple motivating example. Consider the following performative prediction problem.

Example 5. Let individuals and decisions be encoded as signed booleans \( \mathcal{X} = \{ \pm 1 \} \) and \( \hat{Y} = \{ \pm 1 \} \), assuming \( D \) is uniform over \( \mathcal{X} \). Suppose that Nature’s outcome distribution over \( Y = \{ 0, 1 \} \) is governed by the conditional probability function

\[
p^*(x, \hat{y}) = \frac{1}{2} + \beta x \hat{y}
\]

for any \( 0 < \beta < 1/2 \). Consider the goal of learning an \((L, \mathcal{H})\)-performative omnipredictor for the following collection of losses and hypotheses:

- \( L = \{ \ell_0, \ell_1 \} \) contains two opposing steering losses, which steer outcomes towards 0 and 1, respectively.
  \[
  \ell_1(x, \hat{y}, y^*) = 1 - y^* \\
  \ell_0(x, \hat{y}, y^*) = y^*
  \]

- \( \mathcal{H} = \{ h_+, h_- \} \) contains two decision rules over \( \mathcal{X} \) that either returns \( x \) or its negation.
  \[
  h_+(x) = x \\
  h_-(x) = -x
  \]

We begin by considering some naive attempts to achieve the goal of performative omniprediction. Note that \( \mathcal{H} \) actually contains an optimal decision rule for each loss in \( L \). In particular, for losses that steer to 0 versus 1, the optimal decisions minimize (or maximize) the probability that \( y^* = 1 \). The decision rule \( h_+ \) maximizes the probability \( p^*(x, h_+(x)) = 1/2 + \beta \) for all \( x \), whereas \( h_- \) minimizes the probability \( p^*(x, h_-(x)) = 1/2 - \beta \). To obtain the omniprediction guarantee, then, we must learn a probability function that encodes the best decision under \( \ell_0 \) and \( \ell_1 \).

As such, a natural approach would be to fit a function \( p : \mathcal{X} \times \hat{Y} \rightarrow [0, 1] \) that approximates the underlying probability \( p^* \), which we can then post-process for a loss \( \ell \) as in Equation 2. To fit \( p \), we can simply do supervised learning directly over triples \((x, \hat{y}, y^*)\), where \( \hat{y} \) is chosen uniformly at random. We argue that without specification (realizability) assumptions, this approach also fails. Consider, for instance, fitting \( p \) using logistic regression

\[
p(x, \hat{y}) = \frac{1}{1 + \exp(-(ax + b\hat{y} + c))},
\]
where \( a, b, c \) are parameters of the model. In our example, when we select \( \hat{y} \) at random, the outcome \( y^* \) is uncorrelated with each of \( x \) and \( \hat{y} \) on their own. Thus, the optimal setting of these parameters is \( a = b = c = 0 \). Consequently, the logistic model is a constant: \( p(x, \hat{y}) = 1/2 \) for all \( x \in X \) and \( \hat{y} \in \hat{Y} \). Such constant predictions are completely uninformative. Clearly, they cannot suggest the optimal decision rule for any loss, let alone every loss in our collection. The negative result here, follows because the model class for \( p \) was misspecified to fit \( p^* \). In this example, of course, we simply need to run regression with quadratic terms to be well-specified. However, without any assumptions about the complexity of \( p^* \), we cannot rely on approaches that require specifying Nature’s model exactly, which might have unbounded complexity.

### 1.3.1 Performative Outcome Indistinguishability

Recently, [5] introduced the notion of Outcome Indistinguishability (OI) as a new solution concept for supervised learning. In contrast to the traditional framing of learning through loss minimization, OI defines the goal of learning through the lens of indistinguishability. In this view, a predictive model should provide outcomes that cannot be distinguished from true outcomes from Nature. In the world of supervised learning, OI and the closely-related notion of multicalibration [12] have seen broad application, including in deriving supervised omnipredictors [9].

Towards our goal of performative omniprediction, we adapt the paradigm of learning via outcome indistinguishability to the outcome performative setting. In particular, we adapt the loss outcome indistinguishability approach from a recent work of [8]. Intuitively, we say a predictor \( \tilde{p} : X \times \hat{Y} \rightarrow [0, 1] \) is performative outcome indistinguishable if outcomes drawn according to the model \( \hat{y} \sim \tilde{p}(x, h(x)) \) are indistinguishable from Nature’s outcomes \( y^* \sim p^*(x, h(x)) \) under the distribution induced by a decision rule \( h \). To make this notion precise, we need to specify what we mean by indistinguishability and pin down the decision rules we care to reason about.

In particular, to encode omniprediction through performative OI, our goal will be to devise a set of tests of a predictor \( \tilde{p} \) that – if passed – guarantee for every loss \( \ell \in \mathcal{L} \), the decision rule \( \hat{f} \) is as good as any \( h \in \mathcal{H} \). Formally, we start by building a class of tests from a collection of losses \( \mathcal{L} \) and a hypothesis class \( \mathcal{H} \).

**Definition 6 (Performative OI).** For an input distribution \( \mathcal{D} \), collection of losses \( \mathcal{L} \), hypothesis class \( \mathcal{H} \), and \( \varepsilon \geq 0 \), a predictor \( \tilde{p} : X \times \hat{Y} \rightarrow [0, 1] \) is \((\mathcal{L}, \mathcal{H}, \varepsilon)\)-performative outcome indistinguishable (POI) over \( \mathcal{D} \) if for all \( \ell \in \mathcal{L} \) and all \( h \in \mathcal{H} \),

\[
\mathbb{E}_{y^* \sim p^*(x, h(x))} [\ell(x, h(x), y^*)] \approx \varepsilon \mathbb{E}_{\hat{y} \sim \tilde{p}(x, h(x))} [\ell(x, h(x), \hat{y})].
\]

We emphasize how this performative OI condition is a natural notion of outcome indistinguishability. In particular, we note where the predictor \( \tilde{p} \) occurs in the POI conditions: it is only used to sample outcomes according to \( \tilde{p} \) in the modeled world. Importantly, the decisions \( h(x) \) are used in the sampling of \( \hat{y} \). Using \( h(x) \) as the decision associated with \( x \) to sample each outcome \( y \) (under Nature and the model) ensures that \( \tilde{p} \) encodes a reliable estimate of the loss \( \ell \) if the decision rule \( h \) is deployed. Performative OI ensures that \( \tilde{p} \) “knows” these values for each loss in the collection \( \ell \in \mathcal{L} \) and each hypothesis in our class \( h \in \mathcal{H} \).

---

3 More concretely, since \( \mathbb{E}[xy] = \mathbb{E}[yx^*] = 0 \), one can check that \( a = b = c = 0 \) solves the first-order optimality conditions for the logistic regression objective \( \mathbb{E}_{x, y, y^*} [y \log \sigma(ax + by + c) + (1 - y) \log(1 - \sigma(ax + by + c))] \) for \( \sigma(z) = 1/(1 + \exp(-z)) \).

4 Throughout, we use the notational shorthand \( A \approx_s B \) to denote that \( A \in [B - s, B + s] \).
While this OI condition ensures that \( \tilde{p} \) captures the behavior of the hypotheses in \( \mathcal{H} \), it says nothing about the losses under its own decision rules \( \tilde{f} \). Reasoning about these decision rules is essential for performative omniprediction. As such, we introduce an additional OI condition, which we call performative decision OI, to ensure OI under the decision rules suggested by \( \tilde{p} \).

**Definition 7 (Performative Decision OI).** For an input distribution \( \mathcal{D} \), collection of loss functions \( \mathcal{L} \), and \( \varepsilon \geq 0 \), a predictor \( \tilde{p} : \mathcal{X} \times \hat{\mathcal{Y}} \rightarrow [0,1] \) is \((\mathcal{L}, \varepsilon)\)-performative decision outcome indistinguishable (DOI) over \( \mathcal{D} \) if for all \( \ell \in \mathcal{L} \),

\[
\mathbb{E}_{x \sim \mathcal{D}} [\ell(x, \tilde{f}_\ell(x), \tilde{y}^*)] \approx \varepsilon \mathbb{E}_{x \sim \mathcal{D}} [\ell(x, \tilde{f}_\ell(x), \tilde{y})].
\]

Here, \( \tilde{p} \) is still used to sample outcomes \( \tilde{y} \), but is also used to determine the decision rule \( \tilde{f}_\ell \), for each \( \ell \in \mathcal{L} \). Syntactically, changing from \( h \in \mathcal{H} \) to \( \tilde{f}_\ell \) is a small change, but it has significant impacts on the nature of the performative DOI condition – both in terms of its costs and the strength of its guarantees. Critically, for a given loss \( \ell \), the decision rule \( \tilde{f}_\ell \) is (by definition) optimal for outcomes sampled from \( \tilde{y} \sim \tilde{p}(x, \tilde{f}_\ell(x)) \). As such, indistinguishability is a powerful tool here: if the losses are indistinguishable on modeled outcomes (where \( \tilde{f}_\ell \) is optimal) and on Nature’s outcomes, then \( \tilde{f}_\ell \) should be optimal for Nature.

We formalize this intuition, demonstrating that performative OI and decision OI suffice to establish omniprediction. Consider the loss \( \ell \in \mathcal{L} \) obtained by any \( h \in \mathcal{H} \) on true outcomes. We show that the loss of \( \tilde{f}_\ell \) is upper bounded by that of \( h \).

**Proposition 8 (Informal).** If a predictor \( \tilde{p} \) is \((\mathcal{L}, \mathcal{H})\)-POI and \( \mathcal{L}\)-DOI, then \( \tilde{p} \) is an \((\mathcal{L}, \mathcal{H})\)-performative omnipredictor.

**Proof sketch.** Once the appropriate OI conditions are written down, deriving performative omniprediction is almost immediate.

\[
\mathbb{E}_{x \sim \mathcal{D}} [\ell(x, \tilde{f}_\ell(x), \tilde{y}^*)] \approx \varepsilon \mathbb{E}_{x \sim \mathcal{D}} [\ell(x, \tilde{f}_\ell(x), \tilde{y})] \\
\leq \mathbb{E}_{x \sim \mathcal{D}} [\ell(x, h(x), \tilde{y})] \approx \mathbb{E}_{y^* \sim \tilde{p}^*} [\ell(x, h(x), y^*)]
\]

The first equality follows by \( \mathcal{L}\)-DOI, the second equality follows by \((\mathcal{L}, \mathcal{H})\)-POI, and the middle inequality follows by the fact that \( \tilde{f}_\ell \) is optimal over modeled outcomes.

In other words, we have managed to reduce the task of learning performative omnipredictors to learning models satisfying performative OI conditions. Clearly, Nature’s model \( p^* \) is “indistinguishable” from Nature (similarly, it is clear that \( p^* \) is a performative omnipredictor), but the question remains whether there exist *efficient* predictors \( \tilde{p} \) that satisfy the performative OI conditions. Thus, we turn our attention to learning OI predictors under outcome performativity.

### 1.3.2 Learning Outcome Performative Predictors

Despite essential differences in the notions of OI in the supervised and performative settings, we show that many of the algorithmic techniques that have become standard in the literature on multicalibration and OI can be adapted to work in the outcome performative setting. In particular, we demonstrate that, quite generically, learning performative OI models reduces to auditing for distinguishability. Concretely, if there exists a loss \( \ell \in \mathcal{L} \) and hypothesis \( h \in \mathcal{H} \cup \{ \tilde{f}_\ell \} \), such that the performative OI conditions are violated for a predictor \( \tilde{p} \), we can...
use these “distinguishers” to update the model to address the violation. This observation immediately suggests using a boosting algorithm, in the vein of [12], to learn performative OI predictors. Provided that updating based on an \( \varepsilon \)-violation makes significant “progress” towards satisfying performative OI, then the number of auditing steps \( T \leq O(1/\varepsilon^2) \) will be bounded.

While this learning paradigm of “audit, then update” is intuitive, there are nontrivial challenges in maintaining the efficiency of the learned predictors. Consider, for instance, the performative decision OI constraint. As highlighted above, the DOI constraints require that we reason about the optimal decision rule according to \( \tilde{p} \). In particular, to update based on a violation of DOI on loss \( \ell \), we need to incorporate a copy of the function \( \tilde{f}_\ell(\cdot) \). This decision rule, however, is a function of \( \tilde{p}(\cdot, \hat{y}) \) for every \( \hat{y} \in \hat{Y} \) (as it requires computing the argmin over \( \hat{Y} \)). Naively, then, it would seem that in every iteration where we update the model based on some \( \tilde{f}_\ell \), we need to make \( |\hat{Y}| \) recursive oracle calls to the existing model.

To avoid this blow-up, we need to choose a more effective representation of the probability function \( \tilde{p}(\cdot, \cdot) \). We observe that, in general, the updates required for the algorithm are sparse in \( \hat{Y} \). As a result of this sparsity, we can save on overall computation by implementing \( \tilde{p} : \mathcal{X} \times \hat{Y} \to [0,1] \) as a map from individuals to vectors of probabilities \( \tilde{q} : \mathcal{X} \to [0,1]^{\hat{Y}} \). Mathematically, there is a bijection between such functions; computationally, however, the representations behave very differently. By increasing the amount of work per update by a factor of \( |\hat{Y}| \), we avoid making \( |\hat{Y}| \) recursive calls. This strategy is reminiscent of an approach [6] used to learn (supervised) OI predictors for outcomes living in a large domain. With this representation in place, an appropriate analysis reveals that the resulting predictors can be implemented in complexity that scales only polynomially in the number of iterations and decisions.

After addressing representation issues, we study sufficient conditions to implement the auditing task efficiently, from a polynomially bounded number of samples. Achieving performative optimality, in general, requires exploration of the consequence of using different decision rules \( h \in \mathcal{H} \), as these decision rules effect the distribution on outcomes. Nevertheless, we show that it suffices for this exploration to be done “offline” via randomized assignment of decisions \( \hat{y} \in \hat{Y} \). In particular, if we collect triples \( \{(x, \hat{y}, y^*)\} \) through a randomized control trial, assigning \( \hat{y} \) uniformly at random for each \( x \in \mathcal{X} \), and observing \( y^* \sim p^*(x, \hat{y}) \), we avoid the need to deploy each \( h \in \mathcal{H} \).

Given access to such RCT data, we give a reduction from the task of auditing for performative OI to the task of supervised learning for the hypothesis class \( \mathcal{H} \). This reduction from auditing to learning is familiar in the OI framework [12, 5], but critically, we go from a performative prediction task to a non-performative task. In all, our reductions show that if we can learn the best decision rule from \( \mathcal{H} \) in a supervised learning setting, then we can learn performative omnipredictors with respect to \( \mathcal{H} \), assuming access to appropriately sampled data.

### 1.3.3 Universal Adaptability under Outcome Performativity

The OI viewpoint enables a similarly straightforward analysis of distributional robustness, via universal adaptability. Universal adaptability is a notion introduced by [19] in the context of statistical estimation. In the original context, a predictor is universally adaptable if it provides an efficient way to estimate statistics across many underlying input distributions.
We translate the notion of universal adaptability to the outcome performative prediction setting. In our context, we parameterize universal adaptability by a class of importance weight functions \( W \subseteq \{ X \to \mathbb{R}_{\geq 0} \} \). For a base input distribution \( D \), we define a corresponding collection of shifted distributions \( D_W \) to be the set of distributions reachable after reweighting by some \( \omega \in W \).

\[
D_W = \{ D_\omega : \omega \in W, \text{supp}(D_\omega) \subseteq \text{supp}(D) \} \quad \forall x \in \text{supp}(D_\omega) : D_\omega(x) = \omega(x) \cdot D(x)
\]

The key observation is that for any hypothesis \( h \), loss function \( \ell \), and importance weight function \( \omega \), the expected loss over \( D_\omega \) is equal to an expected loss over \( D \), for a loss defined in terms of \( \ell \) and \( \omega \).

\[
E_{x \sim D_\omega, \hat{y} \sim \tilde{p}(x, h(x))} [\ell(x, h(x), \hat{y})] = E_{x \sim D, \hat{y} \sim \tilde{p}(x, h(x))} [\ell(x, h(x), \hat{y}) \cdot \omega(x)]
\]

Importantly, this equality relies on the fact that the outcome probability functions \( p^* \) and \( \tilde{p} \) are defined conditional on \( x \) and \( \hat{y} \), and thus are invariant across shifts in the input distribution.

Using the result that indistinguishability implies omniprediction, by simply enforcing that \( \tilde{p} \) satisfy the POI and DOI conditions relative to \( p^* \) under this enriched class of loss functions \( \ell \cdot \omega \), we can neatly ensure that \( \tilde{p} \) is again POI and DOI, not just over \( D \), but over every marginal distribution \( D_\omega \). Consequently, \( \tilde{p} \) must a be an omnipredictor under all of these \( D_\omega \). The simplicity of this analysis attests to the versatility of the OI perspective and the value it provides in domains beyond supervised learning.

1.4 Related Work and Discussion

Our work lies at the intersection of several areas including performative prediction, the outcome indistinguishability and multicalibration literature, as well as other fields studying algorithmic decision-making such as contextual bandits. We briefly discuss how our results relate to previous work within these areas and conclude with some speculation regarding broader implications of our conclusions.

1.4.1 Performative Prediction

The performative prediction framework was introduced by [28] who defined the main solution concepts and analyzed the convergence of repeated risk minimization to *performatively stable* points. A decision rule \( h_{\text{ps}} \) is performatively stable if it is a fixed point of risk minimization,

\[
h_{\text{ps}} \in \arg\min_{h \in H} \mathbb{E}_{(x, y^*) \sim D_{h_{\text{ps}}}} [\ell(x, h(x), y^*)].
\]

Subsequent work by [24, 4, 2, 3] studied stochastic optimization algorithms for finding stable points in a variety of settings. However, these stable solutions need not be performatively *optimal* as per the definition outlined in Equation 1. In fact, [26] proved that performatively stable models can achieve *arbitrarily worse* loss than performative optimal points. This observation motivated the design of algorithms for findings performatively optimal decision rules for a fixed loss \( \ell \)[26, 13, 14, 27, 15]. These algorithms work in the general performative prediction setup where the model \( h \) can affect the *joint* distribution over pairs \( (x, y) \), but make make very strong specification assumptions on how each \( h \) influences the distribution, and restrict to loss functions satisfying smoothness and strong convexity.
In contrast, our results rely only on the outcome performativity assumption and mild boundedness assumptions. The recent work of [23] has also considered the outcome performativity setting, aiming to understand when performative effects are identifiable from observational data. Short of identifiability, it remains an interesting direction for future research to give learning algorithms for performative omnipredictors from observational data.

Beyond these optimization results, previous work in performative prediction has acknowledged the tension in performative prediction between accurate forecasting and steering. Concretely, [26] discuss how the choice of loss function in performative prediction should balance predictive accuracy with any externalities that arise from the impacts of prediction on the observed distribution. In a different direction, [10] uses performativity as a lens with which to study notions of market power in economics. As part of their analysis, they provide a decomposition of the performatively risk of a classifier into terms that represent forecasting and steering. While we consider how the choice of loss function determines the high-level objective, [10] considers how, even for a fixed loss function, the performatively risk can be decomposed into terms associated with forecasting and steering.

1.4.2 Reinforcement Learning and Contextual Bandits

As discussed in [28], performative prediction, and in particular outcome performativity, can be cast as reinforcement learning (RL) or contextual bandits problems. Individuals $x$ correspond to the contexts, decisions $\hat{y}$ correspond to actions, and the loss $\ell(x, \hat{y}, y^*)$ is captured by the reward $r(x, \hat{y})$. Due to the breadth of their definitions, most ML problems can be written as RL problems.

Still, important issues that arise in outcome performativity – like the tension between forecasting and steering and the desire for omnipredictors – are best seen by focusing on the specific interactions between predictions $\hat{y}$ and outcomes $y$. The variety of losses that can exist for a given outcome are obscured by encapsulating all feedback within an abstract reward function $r(x, \hat{y})$. Moreover, on a technical level, performativity has a richer feedback structure that can be used to design more efficient algorithms as illustrated by [15].

1.4.3 Multicalibration and Outcome Indistinguishability

Originally developed by [12] as a notion of fairness in prediction, multicalibration has seen considerable interest and application in the broader context of supervised learning. At a high-level, multicalibration requires predictions to be calibrated, not just overall, but even when restricting our attention to structured subpopulations. The goal of multicalibration and other related notions of “multi-group” fairness [17, 21, 18, 16] is to ensure that learning occurs within important subpopulations that might otherwise be ignored.

Intuitively, the requirements of multicalibration represent a kind of indistinguishability: calibration requires that the predicted probabilities “look like” real probabilities. [5] formalizes this intuition, introducing the notion of Outcome Indistinguishability, which generalizes multicalibration. They show tight computational equivalences between multi-group fairness notions and variants of OI. Subsequently, OI and multicalibration have been applied in diverse contexts beyond fairness, such as distributional robustness through universal adaptability [19] and omniprediction [9].
1.4.4 Omniprediction

Our work draws inspiration from the work on (supervised) omnipredictors. Even in the supervised learning setting, the existence of efficient omnipredictors is not at all obvious. The main result of [9] demonstrates the feasibility of omnipredictors over any hypothesis class $\mathcal{H}$, for the class $\mathcal{L}_{\text{cvx}}$ of all convex and Lipschitz loss functions. This sweeping result follows by showing that a $\mathcal{H}$-multicalibrated predictor is an $(\mathcal{L}_{\text{cvx}}, \mathcal{H})$-omnipredictor.

On a technical level, our analysis is most closely related to concurrent work by [8] that studies omniprediction in supervised learning through the lens of outcome indistinguishability. By the equivalence of OI with multicalibration, it has been clear since the work of [9] that OI captures loss minimization and omniprediction in the context of supervised learning, albeit indirectly. [8] revisits the question of omnipredictors, directly through the lens of OI, studying a refined notion, which they call loss outcome indistinguishability. Indeed, our proof that Performative OI and Performative Decision OI imply Performative Omniprediction follows a strategy laid out to obtain supervised omnipredictors from loss OI.

While syntactically similar to prior formulations of OI, our notion of performative OI is the first to consider outcome indistinguishability for non-supervised learning distributions. Our objective, in this work, was to derive a notion of performative OI sufficient to imply performative omniprediction. No doubt, further generalizations of the original OI hierarchy [5] to the (outcome) performative setting – and beyond – may prove useful.

1.5 Overview of Full Manuscript

To read the complete set of results and proofs, please see the full version of the paper [20]. We limit ourselves to providing a brief overview of the remaining sections. In Section 2 of the full version we set up outcome performativity formally and establish notation used throughout the manuscript. In Section 3, we define performative omniprediction and performative outcome indistinguishability. There, we show how appropriate performative OI conditions suffice to obtain omniprediction. In Section 4, we establish the universal adaptability properties of performative omnipredictors. In Section 5, we give a generic learning algorithm for performative omnipredictors, demonstrating concrete instantiations of the algorithm using randomized control trial data. Finally, in Section 6, we discuss notions of multicalibration in the context of performative prediction. We speculate that some notions translate to the performative setting naturally, yielding efficient approaches to performative OI, while other notions seem to resist efficient translation.

References

Making Decisions Under Outcome Performativity


14 Zachary Izzo, James Zou, and Lexing Ying. How to learn when data gradually reacts to your model. In International Conference on Artificial Intelligence and Statistics, pages 3998–4035. PMLR, 2022.


