# On the Actionability of Outcome Prediction

**Lydia T. Liu**[1, 2], **Solon Barocas**[1, 3], **Jon Kleinberg**[1], **Karen Levy**[1]
[1]Cornell University
[2]Princeton University
[3]Microsoft Research
{lydiatliu, sbarocas, kleinberg, karen.levy}@cornell.edu

## Abstract

Predicting future outcomes is a prevalent application of machine learning in social impact domains. Examples range from predicting student success in education to predicting disease risk in healthcare. Practitioners recognize that the ultimate goal is not just to predict but to act effectively. Increasing evidence suggests that relying on outcome predictions for downstream interventions may not have desired results.

In most domains there exists a multitude of possible interventions for each individual, making the challenge of taking effective action more acute. Even when causal mechanisms connecting the individual's latent states to outcomes are well understood, in any given instance (a specific student or patient), practitioners still need to infer—from budgeted measurements of latent states—which of many possible interventions will be most effective for this individual. With this in mind, we ask: when are accurate predictors of outcomes helpful for identifying the most suitable intervention?

Through a simple model encompassing actions, latent states, and measurements, we demonstrate that pure outcome prediction rarely results in the most effective policy for taking actions, even when combined with other measurements. We find that except in cases where there is a single decisive action for improving the outcome, outcome prediction never maximizes "action value", the utility of taking actions. Making measurements of actionable latent states, where specific actions lead to desired outcomes, may considerably enhance the action value compared to outcome prediction, and the degree of improvement depends on action costs and the outcome model. This analysis emphasizes the need to go beyond generic outcome prediction in interventional settings by incorporating knowledge of plausible actions and latent states.

## Introduction

Artificial intelligence has been used for impact in variety of societal domains, from education to healthcare. While many of its applications have focused on *prediction*, such as that of educational outcomes (Tamhane et al. 2014; Lakkaraju et al. 2015; Xu et al. 2017) and medical incidents and risk (Hosseinzadeh et al. 2013; Ma et al. 2018; Ballinger et al. 2018; Optum 2024), practitioners and researchers invariably encounter the question of how to use these predictions for interventions to improve the outcomes that they care about.

Consider the example of predicting student academic performance at the secondary level. In most cases, the goal of building such predictors is to improve the relevant educational *outcome*, academic performance. However, a prediction of a student's future academic performance alone does not improve academic performance unless there is an intervening *action*, such as providing additional tutoring or financial support. Students who lack the necessary academic prerequisites may need additional tutoring rather than financial support to improve their performance, whereas students who lack the time to complete course work because they are working multiple jobs may need financial aid rather than to be referred for additional tutoring. Therefore, the success of any action depends on a student's *latent state* (so called as we do not know *a priori* whether the student lacks prerequisites or income). A school official may take *measurements*, such as diagnostic tests, past grades, or a survey of income, to obtain information about students' latent states. But these measurements can be costly, requiring time and labor.

The key question, therefore, is: what should the school official measure in order to best *predict* the student's future academic performance, and what should they measure in order to best *improve* it? Further, when are these the same measurements, and when are they different?

Many applications of ML/AI for social impact focus on solving prediction problems (Kleinberg et al. 2015; Athey 2017) and maximizing prediction accuracy for future outcomes. Although risk prediction has become ubiquitous in education and other domains, its effectiveness for improving outcomes has been called into question. A recent empirical qualitative study by Liu et al. (2023) on machine learning applications in education found a significant gap between predictions and beneficent interventions.

> *"You don't improve things by predicting them better."*
> *- Education researcher on the value of predicting academic risk* (Liu et al. 2023)

The study of interventions has been fundamental in the social sciences, statistics, and theoretical computer science (Rosenbaum and Rubin 1983; Pearl 1995; Rubin 2005; Peters, Janzing, and Schölkopf 2017; Hofman et al. 2021). The set of techniques and applications for causal inference and analysis is vast, mostly notably including program evaluation and randomized controlled trials (Stephenson and Imrie

1998; Deaton and Cartwright 2018), observational studies (Rosenbaum, Rosenbaum, and Briskman 2010) using modern ML techniques (Athey and Imbens 2016), adaptive trial designs (Collins, Murphy, and Strecher 2007; Montoya et al. 2022), individual treatment effect and counterfactual inference (Shalit, Johansson, and Sontag 2017; Lei and Candès 2021; Bynum, Loftus, and Stoyanovich 2023). Prior work informed by causal inference has discussed the gap between predictions and decisions (Athey 2017), and how the use of prediction in these cases is predicated upon critical causal assumptions (Prosperi et al. 2020; Lundberg, Brand, and Jeon 2022). In the education domain, despite the prevalence of RCTs and causal analysis on the population impact of interventions (Cook et al. 2014; Yeager et al. 2019), instance-level targeting and decision making in schools are still often driven by risk scores that predict academic outcomes without incorporating knowledge of plausible interventions (Bruce et al. 2011; Knowles 2015; Perdomo et al. 2023).

The current work is interested in a question that is at the intersection of the pure prediction and the causal inference/interventional paradigm: when is outcome prediction helpful for interventions at the instance level? Given the ubiquity of predictive tools and significant data infrastructure built around prediction, there is a need to better understand the limits of predictions when applying them to interventional settings. This work acknowledges the key role of causal inference, while studying a problem at a different scope—that of instance-level predictive intervention, e.g. what helps this patient, what helps this student, assuming that a model of causal effects is available. Unlike in the estimation of heterogenous treatment effects (see e.g. Imai and Ratkovic 2013; Athey and Imbens 2016), where covariates are assumed to be given, here we investigate the choice of covariates—what to measure, under a constrained budget, in order to predict or intervene. This line of questioning is also related to the theory of *diagnosis* (Reiter 1987; De Kleer and Williams 1987) which has a long history in the AI literature; through the current work we bring the analytical framework of diagnosis to bear on current issues in data-driven prediction and decision making in social systems.

The main contributions of the work are as follows: We formalize the gap between outcome prediction and intervention in a mathematical framework that combines probabilistic modeling, logical formalism, and a theory of action utility. Our model comprises: latent states of individuals, measurements, outcome, and actions to enact change in the latent states (Section ). We then illustrate the actionability of outcome prediction with a simple numerical example in Section . We advance this research agenda in the setting of Boolean functions with a set of theoretical results (Section ): we fully characterize the conditions under which outcome prediction can be considered actionable and show that the optimal measurement for outcome prediction is almost never the optimal measurement for outcome improvement, either when used alone or in combination with other measurements. Rather than prescribe how to best perform interventions (e.g. to improve student academic performance), our goal is to precisely describe when prediction necessarily falls short of intervention goals. In Section , we review further literature in related fields.

## Model

Our model of data-driven decision making comprises four key elements: latent states of individuals, measurements, outcome, and actions to enact change in the latent states. We suppose that an institutional decision maker, whom we refer to as the *planner*, makes measurements for each individual in a population and takes actions based on those measurements for each individual to influence their future outcome. Formally, we have a graphical model with the following random variables, and each individual is an independent random draw from this graphical model.

- *States*. There are $s$ latent states $\mathcal{S} = \{X_1, \cdots, X_s\}$ that are not observed directly, each supported on $\mathcal{D}_{\mathcal{S}}$. Each latent state indicates a factor that influences the individual's outcome, and need not be independent of other states. They have a joint distribution.
- *Outcome*. The outcome of interest is $Y$. The distribution of $Y$ depends on latent states, $Y \sim g(X_1, \cdots, X_s)$ and is supported on $\mathcal{D}_Y$. $g$ is known to the planner. In other words, we assume the planner knows the structural causal model of how states map to the outcome.
- *Measurements*. Planner chooses from $n$ possible measurements $\mathcal{M} := \{M_1, \cdots, M_n\}$. The distribution of $M_i$ depends on States, $M_i \sim f^i(X_1, \cdots, X_s)$ and is supported on $\mathcal{D}_{\mathcal{M}}$. There is a measurement budget of $B$ measurements. The planner observes the realized values of $B$ chosen measurements in order to perform subsequent prediction (Section ) and intervention (Section ) tasks.

To reiterate, each graphical model comprises three types of variables: latent states, measurements, and an outcome (see Figure 1 for an illustration). There is a fourth element of the model, which is *actions*.

*Actions*. After observing the value of measurement(s), the planner takes an action, $a$. Actions change the value of latent states, e.g., the action $a = [X_1 \leftarrow 1]$ changes the value of state $X_1$ to 1.[1] The set of possible actions is denoted $\mathcal{A}$. The cost function of action is $c : \mathcal{A} \to [0, C]$, where $C > 0$. The cost of taking no action, $a = \emptyset$, is 0. This means that the planner can take a (costly) action on behalf of each individual to modify one of their latent states.

In section , we show how the model can be instantiated across three real world problem domains and give examples of the respective states, outcome, measurements and actions.

### Prediction Task and Prediction Value

Consider the case where predicting $Y$ is an end in itself. Then the planner wants to choose a measurement $M \subseteq \mathcal{M}$ such that $|M| \leq B$ and $M$ allows the planner to predict $Y$ with the lowest prediction loss (or error) out of all size $B$ measurement sets. Given a hypothesis class $H$, and prediction loss function $\ell$, the planner constructs an optimal predictor $h^*_M$ given $M$:

$$h^*_M := \operatorname*{argmax}_{h \in H} \mathbb{E}[\ell(h(M), Y)].$$

[1] We use the notation $[X \leftarrow x]$ to denote the *do*-operation that sets the value of variable $X$ to $x$.

For each observed value (or values) of $M$, the optimal predictor $h^*$ outputs a particular prediction of the individual's outcome $Y$. It minimizes prediction loss over the population. We define the *prediction value* of measurement $M$ as:

$$V^{\text{predict}}(M) := -\mathbb{E}[\ell(h^*_M(M), Y)].$$

The higher the prediction value of a measurement, the more informative it is for predicting the outcome, assuming that an optimal predictor is always available to the planner.

## Intervention Task and Action Value

In most cases, the goal of the planner is not simply to predict $Y$, but to take the best action to attain a more favorable outcome $Y$ for the individual. The notion of the "best" action can be made concrete by defining a utility function for actions and outcomes.

Let $Y^a$ denote the outcome variable after an action $a$ has been taken. Action $a$ typically corresponds to a *do*-operation (see e.g. Pearl 2009) on the latent states that changes the distribution of $Y$, e.g., if $a = [X_1 \leftarrow 1]$, then $Y^a = Y^{[X_1 \leftarrow 1]}$ is that new random variable for the outcome under *do*-operation that sets the value of state $X_1$ to 1. Let $u(y)$ denote the utility to the planner of having $Y = y$.

Given any measurement $M \in \mathcal{M}$, the planner constructs an optimal action policy $a^*_M$ to maximize the net utility of taking action:

$$a^*_M := \underset{a:\mathcal{D}_\mathcal{M} \to \mathcal{A}}{\operatorname{argmax}} \mathbb{E}[u(Y^{a(M)})] - \mathbb{E}[u(Y)] - \mathbb{E}[c(a(M))].$$

In words, $a^*_M$ maps any value of $M$ to an action that most improves the expected value of $Y$ conditional on the known value of $M$, taking into account action cost.

We define the *action value* of measurement $M$ as:

$$V^{\text{act}}(M) := \mathbb{E}[u(Y^{a^*(M)})] - \mathbb{E}[u(Y)] - \mathbb{E}[c(a^*(M))].$$

The first term $\mathbb{E}[u(Y^{a^*(M)})]$ is the expected utility under the action policy $a^*$. We may write the first term as

$$\mathbb{E}[u(Y^{a^*(M)})] = \mathbb{E}_M \left[ \mathbb{E}[u(Y^{a^*(m)} \mid M = m] \right].$$

to see that expectation is taken with respect to $Y$ under the *do*-operation (that is, post-action $Y$), as well as with respect to (pre-action) $M$. The second term $\mathbb{E}[u(Y)]$ is the expected utility without taking any action. The third term is the expected cost of the the action policy $a^*$. The higher the action value of a measurement, the most informative it is for taking actions to the improve the outcome in a cost-effective way.

## Motivating Problem Instances

We now discuss motivating real world problems where the model helps to elucidate the different measurements needed for prediction and for intervention. We develop the first example on predicting and improving educational outcomes in some detail, and present the second example on actionable genomics for clinical interventions as a brief sketch.

**Education and student success** Consider the use of student data and machine learning techniques to predict future educational outcomes, such as the student's risk of adverse academic outcomes in secondary school (Lakkaraju et al. 2015) and academic performance in higher education (Bird et al. 2021; Xu et al. 2017; Tamhane et al. 2014). In a critical study by Liu et al. (2023), education researchers that were consulted on the value of making such predictions suggested that the measurements available for making accurate predictions of future educational outcomes (e.g. data on demographic factors, behavioral factors), in the absence of interventions, are not necessarily helpful for selecting interventions to change the outcome.

The model developed in the previous section formally illustrates such concerns and how they arise from the inherent differences between prediction and interventions at the level of measurement.

Suppose the planner is a college official whose mandate is to improve student retention rates. We instantiate the following simplified model of student success:

- *States*. Latent states $X_1, \cdots, X_s$ may include: $X_1$ (whether the student is overworked at job), $X_2$ (whether the student has grasped the academic prerequisites, e.g. calculus), $X_3$ (a demographic feature, e.g. parental education status), etc. These states tend not to be independent, and tend to be only observable via a measurement.
- *Actions.* The corresponding actions are different interventions available to the school official: $[X_1 \leftarrow 1]$ giving financial aid, $[X_2 \leftarrow 1]$ tutoring calculus, etc. There is no corresponding action for $X_3$ as it cannot be modified.
- *Outcome*. The outcome of interest $Y$ is whether the student returns for sophomore year. It is observed at the start of sophomore year. $Y$ is a function of the states, that is, $Y = f(X_1, X_2, X_3, \cdots)$.
- *Measurements*. Some measurements at taken after midterm exams in freshman year. The measurements include $M_1 := X_1$ (student job status), $M_2 := X_2$ (diagnostic calculus test), $M_3 := X_3$ (demographic), $M_4 := f(X_1, X_2, X_3)$ (midterm grades), etc.

In this case, knowing $M_4$ (midterm grades) may be very helpful for predicting $Y$, but it is less helpful for determining which costly action (financial aid or tutoring) should be used to intervene on the student's future retention outcome. In the same vein, $M_1$ and $M_2$ can inform whether the student requires a particular intervention, but without $M_3$, they cannot be used to predict $Y$ as accurately, since $Y$ depends on all three latent states. From an education and testing research perspective, diagnostic tests are different from achievement or proficiency tests (Alderson, Brunfaut, and Harding 2015)—even though $M_2$ and $M_4$ are both test results, the former better informs interventions as it diagnoses specific academic areas that benefit from tutoring.

**Genomics for clinical decisions** This example is taken from Nelson, Keating, and Cambrosio (2013), a study of "actionability" in the context of clinical sequence. Suppose the planner is a hospital with multiple patients to treat. The outcome of interest is health (e.g. the absence of cancer).

Each patient has set of *states* including phenotypes (e.g. whether a patient has a mutated enzyme) and risk factors (e.g. family history of cancer) that together determine their future health outcome. *Measurements* are the genetic sequences of the patient (e.g. whether a patient has a mutation in the anaplastic lymphoma kinase (ALK) gene).

Depending on the type of mutation (e.g. hereditary mutations in pre-symptomatic individuals or non-heritable sporadic mutations), a state may or may not be associated with an *action* that can improve the patient's health outcome: non-heritable mutations in the tumor may be associated with a drug mechanism that that can block the function of the mutated enzyme, whereas gene markers that are associated with future health risks typically cannot be targeted by any particular drug pathway.

## Illustrative Example With Two Latent States

In this section, we work out a simple example of the model to illustrate how the prediction value and action value of measurements can be misaligned. In this instantiation of the model, we assume that all variables are binary. In the motivating problems discussed above, latent states are often adequately modeled as binary variables, e.g., a patient either has a genetic mutation or not. Many outcomes of interest discussed are also naturally binary, such as testing positive for a disease (health outcome), or graduating on time (educational outcome); many real-valued outcomes can also be turned into a binary outcome by applying an appropriate threshold. While the general version of the model as described in Section  does not require variables to be binary-valued, we argue that such models suitably represent certain real world problems of interest.The particular graphical model we analyze in this section is displayed in Figure 1, and described as follows.

- *States*. There are two latent states $X_1, X_2$ distributed as independent Bernoulli random variables with failure rate $p < 0.5$. That is, we have $X_1, X_2 \sim \text{Bernoulli}(1-p)$.
- *Outcome*. The outcome of interest is $Y := X_1 \wedge X_2$, where $\wedge$ denotes the logical and.
- *Measurements*. The space of measurements $\mathcal{M}$ is all Boolean functions of $(X_1, X_2)$. The measurement budget is $B = 1$. For the purposes of this example, we focus on the following 3 measurements:

  $$M_1 := X_1; \quad M_Y := X_1 \wedge X_2; \quad M_{\text{piv}} := X_1 \wedge \neg X_2.$$

  By the symmetry of the example, the other plausible measurements such as $X_2$ and $X_2 \wedge \neg X_1$ follow similar calculations. We call $M_{\text{piv}}$ a *pivotal* measurement, which indicates that a particular state is pivotal for changing the outcome.[2]
- *Actions*. The actions are $\mathcal{A} := \{[X_1 \leftarrow 1], [X_2 \leftarrow 1], \emptyset\}$. The cost of action is fixed for $[X_i \leftarrow 1]$ at $c > 0$.
- *Utility*. Planner's utility from outcome $Y$ is $u(y) = y$.

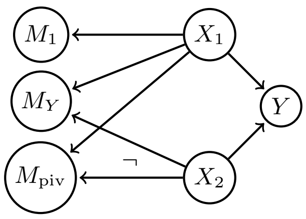

Figure 1: Binary variable model with 2 latent states. Arrows indicate logical addition unless otherwise stated.

### Prediction Value

Suppose the hypothesis class $H$ is any (potentially randomized) Boolean function on $\{0, 1\}$ and we consider the 0-1 loss. For notational brevity, we write $q = 1 - p$ whenever necessary. By elementary calculations, we know:

- The best predictor of $Y$, given $M_1$ is to predict 1 when $M_1 = 1$ and 0 when $M_1$ = 0.
- The best predictor of $Y$, given $M_Y$ is to predict $M_Y$.
- The best predictor of $Y$, given $M_{\text{piv}}$, is
  - If $q^2 > p$: predict 0 when $M_{\text{piv}} = 1$ and predict 1 when $M_{\text{piv}} = 0$;
  - Otherwise: always predict 0. (See footnote[3].)

We may compute the (negative) prediction value of each measurement as follows. This is none other than the expected loss of the respective optimal predictor:

$$\begin{aligned} -V^{\text{predict}}(M_1) &= 1 - \mathbb{P}(X_1 = 1, X_2 = 0) \\ &= 1 - pq, \\ -V^{\text{predict}}(M_Y) &= 1 \\ -V^{\text{predict}}(M_{\text{piv}}) &= \max(1-p, 1-q^2). \end{aligned}$$

Hence, ranking the measurements by prediction value, we have $M_Y \succeq M_1 \succeq M_{\text{piv}}$.

### Action Value

We turn to the intervention task. To compute action values, we analyze the best action policies given each measurement. First consider $M_{\text{piv}}$. In this case, the best action policy is:

- If $M_{\text{piv}} = 1$, we have $X_1 = 1$ and $X_2 = 0$. The best action is $[X_2 \leftarrow 1]$.
- If $M_{\text{piv}} = 0$, the best action depends on $c$ and $p$. If $\mathbb{P}(X_2 \wedge \neg X_1 = 1 \mid M_{\text{piv}} = 0) > c$, the best action is $[X_1 \leftarrow 1]$. Otherwise, the best action is to do nothing.

The key takeaway is that $M_{\text{piv}}$ allows the planner to take the action that is *pivotal* for improving the outcome. This both maximizes the utility gain from successfully improving the outcome, and minimizes the cost of taking actions.

[2] In the full version of the paper at arXiv:2309.04470, we describe examples of real world pivotal measurements , such as "telltale" symptoms of diseases that exclude other conditions and suggest a clear treatment path, and contrast them with non-pivotal versions.

[3] When $M_{\text{piv}} = 1$, $Y$ must be 0, that is $\mathbb{P}(Y = 0 \mid M_{\text{piv}} = 1) = 1$ but in the case when $M_{\text{piv}} = 0$, we have $\mathbb{P}(Y = 1 \mid M_{\text{piv}}) > \mathbb{P}(Y = 0 \mid M_{\text{piv}})$ if and only if $q^2 > p$.

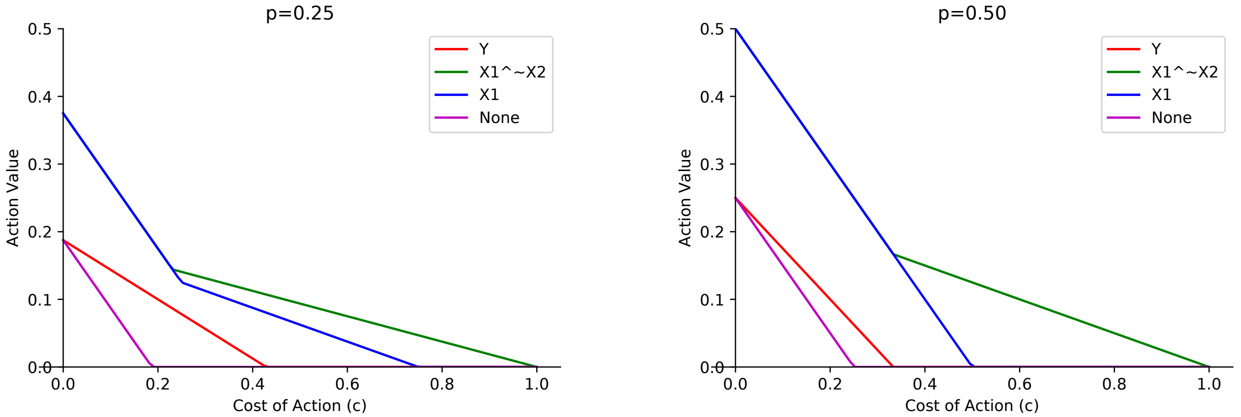

Figure 2: Action value against action cost for 3 measurements: $M_Y$ (Highest prediction value), $M_{\text{piv}}$, and $M_1$. The action value of making no measurements is included as a baseline. Failure rate $p$ is set to $0.25$ in left plot, and to $0.5$ in the right plot.

Performing similar analyses for $M_Y$ and $M_1$, we find that the action value of each measurement is:

$$\begin{aligned}
V^{\text{act}}(M_1) &= \max(0, pq - pc) + \max(0, pq - qc)\\
V^{\text{act}}(M_Y) &= \max(0, pq - (1-q^2)c)\\
V^{\text{act}}(M_{\text{piv}}) &= \max(0, pq(1-c)) + \max(0, pq - (1-pq)c.
\end{aligned}$$

In Figure 2, we plot the action value against cost $c$ for two different $p$ parameters. First we observe that $M_{\text{piv}}$ has the highest action value regardless of action costs.

We also observe that when actions are very costly, the measurement $M_Y$ is no longer helpful in terms of action utility, i.e., $V^{\text{act}}(M_Y) = 0$. $M_Y$ corresponds to perfect knowledge of the outcome $Y$, which is typically not possible in reality. Yet, even under this favorable assumption, we see that knowing the outcome has rather limited utility for effective intervention.

In contrast, $M_{\text{piv}}$ and $M_1$, which as we recall have lower prediction value, help to inform good action policies. When actions are low-cost, all measurements have positive action value. When actions are very low-cost, $M_{\text{piv}}$ and $M_1$ have the same action value and their advantage over $M_Y$ is even greater than when actions are costly.

Ranking the measurements by action value, we have $M_{\text{piv}} \succeq M_1 \succeq M_Y$. In this case, the order is completely reversed from the ranking by prediction value. In the section that follows, we will see that this is an instance of a more general phenomenon.

## Main Results: Boolean Outcome Functions

In this section, we continue with the analysis of outcome models with binary variables and consider a more general setting, where $Y$ is a Boolean function of $s$ States, $X_1, \cdots, X_s$. That is, $Y : \{0,1\}^s \to \{0,1\}$. As noted in the previous section, while this is a simplified setting where the states, and the outcome, can be either 0 or 1 (e.g., favorable or unfavorable), the space of Boolean functions is sufficiently rich to capture wide range of interactions between states and outcome.

Suppose that we can measure $M$, any Boolean function of $X_1, \cdots, X_s$; in other words, $\mathcal{M}$ is the set of all Boolean functions over $\{0,1\}^s$. The set of possible actions is $\mathcal{A} = \bigcup_{i=1,\cdots,s,x\in\{0,1\}}\{[X_1 \leftarrow x]\} \cup \emptyset$. In words, the planner can set the value of any state to 0, or 1, or do nothing.

### Prediction and Action for a Single Measurement

We first consider the case where the measurement budget is $B = 1$. In Proposition 1, we work through an example for symmetric and monotone outcome function $Y$ where the states are i.i.d. Bernoulli random variables. Then in the main result (Theorem 2), we give the sufficient and necessary condition for outcome prediction to have maximum action value for any Boolean $Y$. The condition results in a very constrained outcome model, where a single latent state always improves the outcome (see Definition 1). In other words, outcome prediction never has the optimal action value except in highly degenerate models.

The following illustrative result generalizes the example in Section . Proposition 1 gives an explicit expression for a measurement $M^*$ that has higher action value than the measurement $M_Y$ that perfectly tracks the outcome $Y$. $M^*$ generalizes the pivotal measurement that was introduced in the previous section.

**Proposition 1** (Construction of measurement with high action value)**.** *Suppose $Y$ is a symmetric and monotone Boolean function of $s$ States, $X_1, \cdots, X_s$, which are i.i.d. Bernoulli$(1-p)$ random variables. We can measure $M$, any Boolean function of $X_1, \cdots, X_s$, and take any action $a \in \mathcal{A}$ for a fixed cost $c \in (0,1)$. Then, the following measurement $M^*$ has higher action value than $M^{\text{predict}} = Y$ for any $c \in (0,1)$:*

$$\begin{aligned}
&M^*(a_1, \cdots, a_s) = 1 \iff\\
&Y(a_1, \cdots, a_{i-1}, 1-a_i, a_{i+1}, \cdots, a_s)\\
&-Y(a_1, \cdots, a_{i-1}, a_i, a_{i+1}, \cdots, a_s) = 1,
\end{aligned}$$

*for any $i = 1, \cdots, s$. Moreover, the inequality is strict for all but univariate $Y$.*

The proof [4] proceeds by deriving explicit algebraic expressions for action values in terms of model parameters.

The remaining goal of this section is to generalize the above proposition to arbitrary Boolean outcome function $Y$. To do so, we introduce two new definitions.

**Definition 1.** *An outcome $Y$ is **fully improvable** if for any $x_1, \cdots, x_s$ where $\mathbb{P}(X_1 = x_1, \cdots, X_s = x_s) > 0$ and $Y(x_1, \cdots, x_s) = 0$, there exists $i, x$, s.t.*

$$\mathbb{P}(Y^{[X_i \leftarrow x]} = 1 \mid X_1 = x_1, \cdots, X_s = x_s) = 1.$$

**Definition 2.** *The action $[X_{i^*} \leftarrow x]$ is **sufficient** for improving $Y$ if $\forall x_1, \cdots, x_s$ s.t. $\mathbb{P}(X_1 = x_1, \cdots, X_s = x_s) > 0$,*

$$\mathbb{P}(Y^{[X_{i^*} \leftarrow x]} = 1 \mid X_1 = x_1, \cdots, X_s = x_s) = 1.$$

Full improvability is a strong condition on the outcome $Y$ which states that there always exists a single action on the latent states that improves $Y$ almost surely. This single action can in general depend on the latent states. Full improvability is already a restrictive condition: threshold functions, $\mathbf{1}\{\sum_{i=1}^{s} X_i \geq k\}$, are not fully improvable for $k > 1$.

The existence of a sufficient action is an even stronger condition which states that the same single action improves $Y$ almost surely across all realizations of the latent states. We note that having a sufficient action indicates that the outcome $Y$ is fully improvable, but the former does not necessarily imply the latter. For example, the parity function of $s$ Boolean variables is fully improvable, but it does not have a sufficient action. In words, having a sufficient action means that there's a single action that improves $Y$ whenever $Y = 0$ regardless of the configuration of the latent states; that is, one treatment helps all equally. We now show that predicting the outcome is optimal for taking actions if and only if the strong and likely unrealistic condition—of having a sufficient action—holds.

**Theorem 2** (Outcome prediction and maximum action value)**.** *Let $Y(X_1, \cdots, X_s)$ be an $s$-dimensional Boolean function such that $Y \not\equiv 0$. If $Y$ does not have a sufficient action, there exists $M(X_1, \cdots, X_s)$ such that $V^{\text{act}}(M) > V^{\text{act}}(Y)$ for $c < 1$. If $Y$ has a sufficient action $[X_{i^*} \leftarrow x]$, then $V^{\text{act}}(Y)$ is maximal for any cost $c$.*

*Proof.* We prove the first direction, that is, we assume $Y$ does not have a sufficient action. Suppose the best action given $Y = 0$ is $do(X_1 \leftarrow x)$, WLOG. Since the best action given $Y = 1$ is $\emptyset$, the action value of $Y$ is

$$V^{\text{act}}(Y) = \max(0, \mathbb{P}(Y = 0 \cap Y^{[X_{i^*} \leftarrow x]} = 1) - \mathbb{P}(Y = 0) \cdot c)$$

Let $M$ be s.t. $\{M = 0\} = \{Y = 0\} \cap \{Y^{[X_{i^*} \leftarrow x]} = 1\}$. The action value of $M$ is

$$V^{\text{act}}(M) \geq \max(0, \mathbb{P}(Y = 0 \cap Y^{[X_{i^*} \leftarrow x]} = 1) \cdot (1 - c)).$$

By assumption, we have that $\mathbb{P}(Y = 0) > \mathbb{P}(M = 0)$. Therefore, for any $c < 1$, we have $V^{\text{act}}(M) > V^{\text{act}}(Y)$.

Now, for the other direction. Recall that the action value of $Y$ is $\mathbb{P}(Y = 0) \cdot (1 - c)$. Consider some measurement $M$ that is $(X_1, \cdots, X_s)$-measurable, and suppose the optimal action policy given $M$ is:

$$a(M) = \begin{cases} [X_i \leftarrow x_i] & \text{if } M = 0 \\ [X_j \leftarrow x_j] & \text{if } M = 1 \end{cases}.$$

Then the action value of $M$ is $V^{\text{act}}(M)$

$$\begin{aligned}
&= \max(0, \mathbb{P}(Y = 0, M = 0, Y^{[X_i \leftarrow x_i]} = 1) \\
&\qquad - \mathbb{P}(M = 0) \cdot c) \\
&\quad + \max(0, \mathbb{P}(Y = 0, M = 1, Y^{[X_j \leftarrow x_j]} = 1) \\
&\qquad - \mathbb{P}(M = 1) \cdot c) \\
&\leq \max(0, \mathbb{P}(Y = 0, M = 0) - \mathbb{P}(M = 0, Y = 0) \cdot c) \\
&\quad + \max(0, \mathbb{P}(Y = 0, M = 1) - \mathbb{P}(M = 1, Y = 0) \cdot c) \\
&= \mathbb{P}(Y = 0, M = 0) \cdot (1 - c) \\
&\quad + \mathbb{P}(Y = 0, M = 1) \cdot (1 - c) \\
&= \mathbb{P}(Y = 0) \cdot (1 - c).
\end{aligned}$$

This shows that $Y$ has the maximal action value among all measurements. □

The interested reader may refer to the full version of the paper for an illustration of the proof idea.

## Prediction and Action for a Measurement Set

In this section, we consider measurement sets of size $B > 1$ and we prove a generalization of the second implication in Theorem 2—that $Y$ is typically not part of a set of measurements that together maximizes action value.

As we turn our consideration from single measurements to measurement sets, information that is conveyed from certain measurements may become redundant. On the question of whether $Y$ is an actionable measurement when used in combination with other measurements, we would therefore like to focus on measurement sets where $Y$ is **non-redundant**, defined as follows.

**Definition 3.** *Consider $Y \cup S$, a size-$B$ measurement set containing $Y$ for $S$ such that $|S| = B - 1$ and $S \subseteq \mathcal{M}$. We say that $Y$ is **non-redundant** with respect to $S$ if there exists $s \in \{0, 1\}^{B-1}$ such that the best action when $Y = 0, S = s$ is not $\emptyset$. We call the set of such $s$ the $Y$**-relevant set** with respect to $S$:*

$$\{s \in \{0, 1\}^{B-1} : a^*(Y = 0, S = s) \neq \emptyset\}.$$

Note that the best action when $Y = 1$ is always $\emptyset$, so by Definition 3, the non-redundant set of $Y$ with respect to $S$ is where optimal action is dependent on $Y$ conditioning on $S$. We also extend the notion of sufficient action to subsets of the probability space.

**Definition 4.** *For any $\mathcal{S}$-measurable set $\mathcal{F}$, the action $[X_{i^*} \leftarrow x]$ is **sufficient for improving** $Y$ **on** $\mathcal{F}$ if $\forall x_1, \cdots, x_s$ s.t. $\mathbb{P}(X_1 = x_1, \cdots, X_s = x_s \mid \mathcal{F}) > 0$,*

$$\mathbb{P}(Y^{[X_{i^*} \leftarrow x]} = 1 \mid X_1 = x_1, \cdots, X_s = x_s) = 1.$$

In the following theorem, we show that $Y$ cannot be a element of an optimal measurement set, unless $Y$ is a redundant measurement, or a strong condition is satisfied: that $Y$ has a sufficient action whenever it is non-redundant.

[4] The proof can be found in the full version of the paper at arXiv:2309.04470.

**Theorem 3** (Action value of measurement sets containing the outcome can be improved)**.** *Let $Y(X_1, \cdots, X_s)$ be an $s$-dimensional Boolean function. Consider $Y \cup S$, the size-$B$ measurement set where $Y$ is non-redundant. Suppose there exists $\bar{s} \in \{0,1\}^{B-1}$ in the $Y$-relevant set with respect to $S$ such that $Y$ does not have a sufficient action on $\{S = \bar{s}\}$. Then, there is a measurement $M^*$ such that*

$$V^{\text{act}}(Y \cup S) < V^{\text{act}}(M^* \cup S).$$

*Proof.* By assumption, there exists $\bar{s} \in \{0,1\}^{B-1}$ such that $a^*(Y = 0, S = \bar{s}) = [X_i \leftarrow x]$ and $[X_i \leftarrow x]$ is not a sufficient action for improving $Y$ on $\{S = \bar{s}\}$.

Let $X_{NI}$ denote the set of state values $(x_1, \cdots, x_s)$ where $S = \bar{s}$ and the action $[X_i \leftarrow x]$ does not improve $Y$, that is,

$$\mathbb{P}(Y^{[X_i \leftarrow x]} = 1 \mid X_1 = x_1, \cdots, X_s = x_s) = 0.$$

Since $[X_i \leftarrow x]$ is not a sufficient action on $\{S = \bar{s}\}$, we must have $\mathbb{P}((X_1, \cdots, X_s) \in X_{NI}) > 0$.

Construct a new measurement $M^*$ such that

$$M = \begin{cases} 1 & \text{if } Y = 1 \text{ or } (X_1, \cdots, X_s) \in X_{NI} \\ 0 & \text{o.w.} \end{cases}$$

Compare the best action policy under $Y \cup S$ and $M^* \cup S$. The best action policy changes only on the set $\{(X_1, \cdots, X_s) \in X_{NI}\}$, where the planner now takes no action instead of $[X_i \leftarrow x]$. The action value is therefore improved by $\mathbb{P}((X_1, \cdots, X_s) \in X_{NI}) \cdot c > 0$. We have shown that the measurement set $M^* \cup S$ has strictly higher action value than $Y \cup S$. □

In Theorem 3, recall that we assumed $Y$ does not have a sufficient action on some element of the $Y$-relevant set, which implies and is stronger than the condition that $Y$ does not have a sufficient action overall. The following example shows that this assumption is necessary. Consider a slightly modified outcome model from Section : there are two binary latent states where $Y = X_1 \wedge X_2$, and the marginal distribution of $X_1, X_2$ is Bernoulli but we have $\mathbb{P}(X_1 = X_2 = 0) = 0$. Here, $Y$ does not have a sufficient action and yet $\{Y, X_1 \wedge \neg X_2\}$ is an optimal measurement set (maximum action value among all measurement sets).

In this section, we derived theoretical results for models where the outcome and measurements are Boolean functions of latent states, showing that outcome prediction does not in general maximize action value. We leave the interesting question of extending these results to other settings, such as when the variables are real-valued, to future work.

## Discussion

In this paper, we studied the gap between outcome prediction and intervention in a probabilistic graphical model of outcomes, states, actions and measurements. By distinguishing between the utility of a measurement for accurate prediction and for effective intervention, we show that outcome prediction almost never leads to an optimal measurement or an optimal measurement set for interventions. Our result is framed theoretically at a general level, to provide a language for reasoning about predictions and actions beyond the specifics of any one domain. We now elaborate on further connections with related work, and close with a discussion of open questions.

**Heterogeneous Causal Effects and Policy Evaluation** Athey and Imbens (2016); Shalit, Johansson, and Sontag (2017); Wager and Athey (2018) have examined the estimation of heterogeneous treatment effects from observational data. Furthermore, the area of off-policy learning and optimization (Manski 2004; Zhao et al. 2012; Dudík et al. 2014; Kallus and Zhou 2018; Athey and Wager 2021) studies average causal outcomes under personalized treatment assignment policies, exemplified in studies focusing on job training interventions (Kitagawa and Tetenov 2018; Knaus, Lechner, and Strittmatter 2022). The framework of policy optimization is an alternative framework for algorithmic decision making that precludes the need for outcome predictors and human-in-the-loop decision making; it requires data about treated and untreated outcomes, and the treatment policy under which data was collected. Typically, the estimation of heterogenous treatment effects is limited to scenarios where only a single treatment (either discrete or continuous) is considered, without delving into the problem of diagnosing multiple causal factors. Going beyond randomized controlled trials, adaptive interventions involving multiple assignment strategies has become increasingly popular in the clinical application domain (Collins, Murphy, and Strecher 2007; Montoya et al. 2022). Though we have similar goals of finding optimal personalized treatment assignments—called an "action policy" in the current work—personalization in this line of work depends on given covariates, whereas the current model examines the choice of what covariates to measure under a measurement budget.

**Recourse and Strategic Action in Machine Learning** A rich literature has developed over recent years on the topic of *recourse*—that is, how individuals subject to an adverse decision based on a machine learning model might change their feature values to achieve a more favorable decision in the future (Ustun, Spangher, and Liu 2019; Verma et al. 2020; Ross, Lakkaraju, and Bastani 2021; Karimi et al. 2022). The research emphasizes the need for explanations that highlight mutable and more easily changeable features to guide individual action (Joshi et al. 2019; Barocas, Selbst, and Raghavan 2020; Karimi, Schölkopf, and Valera 2021). While this work shares a common motivation with the present paper—to help individuals to achieve desired outcomes rather than just predict likely outcomes—it differs in two crucial ways. First, the work on recourse is specifically focused on the actions that can be taken by decision subjects, whereas we are concerned with the actions available to a social planner who is generally seeking to achieve positive societal impact. Secondly, while recourse focuses on altering the decisions output by a machine learning model, we are concerned with actions that affect the likelihood of the actual outcome of interest, not merely a model's predictions.

The growing body of research on *strategic classification* aims to assess how decisions subjects might adapt their behavior in light of a machine learning model making decisions (Brückner and Scheffer 2011; Hardt et al. 2016;

Kleinberg and Raghavan 2019; Hu, Immorlica, and Vaughan 2019; Milli et al. 2019; Liu et al. 2020). This line of work explores the concept of gaming, where individuals manipulate input features to improve model predictions without necessarily affecting the underlying property. Prior work has demonstrated that limiting strategic behavior along these lines requires causal modeling (Miller, Milli, and Hardt 2020; Shavit, Edelman, and Axelrod 2020). This area of work focuses on designing the right incentives within an ML model to cause decision subjects to behave as the social planner might like them to behave. In contrast, the current work is focused on designing the measurements that help the social planner achieve its interventional goals directly.

**Open Questions** This theoretical investigation contributes to the discourse around the actionability of risk prediction in the education domain and beyond (Liu et al. 2023) and argues that machine learning practitioners should reconsider risk prediction that captures only a single outcome when their goals are interventional. That being said, our model assumes that the planner knows the structural causal model of how states map to the outcome, making it possible to maximize the action value of measurements. Recovering the structural causal model is difficult to do in full generality. In some cases, however, experts might have domain knowledge of some causal connections between nodes, e.g. clinicians may know about particular drug pathways that connects a drug (action) to a patient's outcome (as discussed in Section ). Leveraging known structural causal models, our study identifies situations where measuring latent states is preferable to outcome prediction. The characterization of the action value of measurements in partially known outcome models is left as an open question.

Relatedly, our findings suggest the need to invest in methods for determining both the structural causal model and measuring latent states, especially for practitioners prioritizing action value over prediction value. Recent work has highlighted the fraught multiplicity and temporality of predicted risk notions (Saxena et al. 2023), as well as other shortcomings of optimizing machine learning models to predict future outcomes (Wang et al. 2022). Future research could explore viable alternatives, such as operationalizing more actionable measurements or integrating knowledge of interventions into predictive models that are in production, in addition to recognizing conditions under which one might decide not to implement prediction at all (Garcia et al. 2020).

While existing studies (Perdomo et al. 2023; Liu et al. 2023) have considered the question of whether outcome predictions result in better interventional outcomes, such empirical studies are emerging and few. Existing open source machine learning datasets do not typically include data on actions taken and downstream outcomes, against which to evaluate the action value of prediction. The current work has explored this important problem via theoretical analysis, suggesting the need for further empirical inquiry and updated practices for model evaluation.

## Acknowledgments

We would like to acknowledge the support of the John D. and Catherine T. MacArthur Foundation and the Simons Foundation. We would like to thank members of the Artificial Intelligence, Policy and Practice (AIPP) initiative for feedback on an earlier iteration of this work.

## References

Alderson, J. C.; Brunfaut, T.; and Harding, L. 2015. Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, 36(2): 236–260.

Athey, S. 2017. Beyond prediction: Using big data for policy problems. *Science*, 355(6324): 483–485.

Athey, S.; and Imbens, G. 2016. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Athey, S.; and Wager, S. 2021. Policy learning with observational data. *Econometrica*, 89(1): 133–161.

Ballinger, B.; Hsieh, J.; Singh, A.; Sohoni, N.; Wang, J.; Tison, G.; Marcus, G.; Sanchez, J.; Maguire, C.; Olgin, J.; et al. 2018. DeepHeart: semi-supervised sequence learning for cardiovascular risk prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Barocas, S.; Selbst, A. D.; and Raghavan, M. 2020. The hidden assumptions behind counterfactual explanations and principal reasons. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 80–89.

Bird, K. A.; Castleman, B. L.; Mabel, Z.; and Song, Y. 2021. Bringing Transparency to Predictive Analytics: A Systematic Comparison of Predictive Modeling Methods in Higher Education. *AERA Open*, 7: 23328584211037630.

Bruce, M.; Bridgeland, J. M.; Fox, J. H.; and Balfanz, R. 2011. On Track for Success: The Use of Early Warning Indicator and Intervention Systems to Build a Grad Nation. *Civic Enterprises*.

Brückner, M.; and Scheffer, T. 2011. Stackelberg games for adversarial prediction problems. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 547–555.

Bynum, L. E. J.; Loftus, J. R.; and Stoyanovich, J. 2023. Counterfactuals for the Future. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(12): 14144–14152.

Collins, L. M.; Murphy, S. A.; and Strecher, V. 2007. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American journal of preventive medicine*, 32(5): S112–S118.

Cook, P. J.; Dodge, K.; Farkas, G.; Fryer, R. G.; Guryan, J.; Ludwig, J.; Mayer, S.; Pollack, H.; Steinberg, L.; et al. 2014. The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: Results from a randomized experiment in Chicago. Technical report, National Bureau of Economic Research.

De Kleer, J.; and Williams, B. C. 1987. Diagnosing multiple faults. *Artificial intelligence*, 32(1): 97–130.

Deaton, A.; and Cartwright, N. 2018. Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210: 2–21.

Dudík, M.; Erhan, D.; Langford, J.; and Li, L. 2014. Doubly Robust Policy Evaluation and Optimization. *Statistical Science*, 485–511.

Garcia, P.; Sutherland, T.; Cifor, M.; Chan, A. S.; Klein, L.; D'Ignazio, C.; and Salehi, N. 2020. No: Critical Refusal as Feminist Data Practice. In *3rd ACM Conference on Computer-Supported Cooperative Work and Social Computing, CSCW 2020*, 199–202. Association for Computing Machinery.

Hardt, M.; Megiddo, N.; Papadimitriou, C.; and Wootters, M. 2016. Strategic Classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, 111–122. New York, NY, USA: ACM. ISBN 978-1-4503-4057-1.

Hofman, J. M.; Watts, D. J.; Athey, S.; Garip, F.; Griffiths, T. L.; Kleinberg, J.; Margetts, H.; Mullainathan, S.; Salganik, M. J.; Vazire, S.; et al. 2021. Integrating explanation and prediction in computational social science. *Nature*, 595(7866): 181–188.

Hosseinzadeh, A.; Izadi, M.; Verma, A.; Precup, D.; and Buckeridge, D. 2013. Assessing the predictability of hospital readmission using machine learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 27, 1532–1538.

Hu, L.; Immorlica, N.; and Vaughan, J. W. 2019. The Disparate Effects of Strategic Manipulation. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, 259–268. New York, NY, USA: ACM. ISBN 978-1-4503-6125-5.

Imai, K.; and Ratkovic, M. 2013. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1): 443.

Joshi, S.; Koyejo, O.; Vijitbenjaronk, W.; Kim, B.; and Ghosh, J. 2019. Towards realistic individual recourse and actionable explanations in black-box decision making systems. *arXiv preprint arXiv:1907.09615*.

Kallus, N.; and Zhou, A. 2018. Policy evaluation and optimization with continuous treatments. In *International conference on artificial intelligence and statistics*, 1243–1251. PMLR.

Karimi, A.-H.; Barthe, G.; Schölkopf, B.; and Valera, I. 2022. A survey of algorithmic recourse: contrastive explanations and consequential recommendations. *ACM Computing Surveys*, 55(5): 1–29.

Karimi, A.-H.; Schölkopf, B.; and Valera, I. 2021. Algorithmic recourse: from counterfactual explanations to interventions. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 353–362.

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

Kleinberg, J.; Ludwig, J.; Mullainathan, S.; and Obermeyer, Z. 2015. Prediction policy problems. *American Economic Review*, 105(5): 491–95.

Kleinberg, J.; and Raghavan, M. 2019. How Do Classifiers Induce Agents to Invest Effort Strategically? In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, 825–844. New York, NY, USA: ACM. ISBN 978-1-4503-6792-9.

Knaus, M. C.; Lechner, M.; and Strittmatter, A. 2022. Heterogeneous employment effects of job search programs: A machine learning approach. *Journal of Human Resources*, 57(2): 597–636.

Knowles, J. E. 2015. Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin. *Journal of Educational Data Mining*, 7(3): 18–67.

Lakkaraju, H.; Aguiar, E.; Shan, C.; Miller, D.; Bhanpuri, N.; Ghani, R.; and Addison, K. L. 2015. A machine learning framework to identify students at risk of adverse academic outcomes. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 1909–1918.

Lei, L.; and Candès, E. J. 2021. Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(5): 911–938.

Liu, L. T.; Wang, S.; Britton, T.; and Abebe, R. 2023. Reimagining the machine learning life cycle to improve educational outcomes of students. *Proceedings of the National Academy of Sciences*, 120(9): e2204781120.

Liu, L. T.; Wilson, A.; Haghtalab, N.; Kalai, A. T.; Borgs, C.; and Chayes, J. 2020. The disparate equilibria of algorithmic decision making when individuals invest rationally. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 381–391.

Lundberg, I.; Brand, J. E.; and Jeon, N. 2022. Researcher reasoning meets computational capacity: Machine learning for social science. *Social science research*, 108: 102807.

Ma, F.; Gao, J.; Suo, Q.; You, Q.; Zhou, J.; and Zhang, A. 2018. Risk prediction on electronic health records with prior medical knowledge. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1910–1919.

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

Miller, J.; Milli, S.; and Hardt, M. 2020. Strategic classification is causal modeling in disguise. In *International Conference on Machine Learning*, 6917–6926. PMLR.

Milli, S.; Miller, J.; Dragan, A. D.; and Hardt, M. 2019. The Social Cost of Strategic Classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, 230–239. New York, NY, USA: ACM. ISBN 978-1-4503-6125-5.

Montoya, L. M.; Kosorok, M. R.; Geng, E. H.; Schwab, J.; Odeny, T. A.; and Petersen, M. L. 2022. Efficient and Robust Approaches for Analysis of SMARTs: Illustration using the ADAPT-R Trial. *arXiv preprint arXiv:2210.03316*.

Nelson, N. C.; Keating, P.; and Cambrosio, A. 2013. On being "actionable": clinical sequencing and the emerging contours of a regime of genomic medicine in oncology. *New Genetics and Society*, 32(4): 405–428.

Optum. 2024. Health Risk Analytics - Impact Pro. https://www.optum.com/business/health-plans/data-analytics/predict-health-risk.html. Accessed: 23-01-2024.

Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika*, 82(4): 669–688.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. New York, NY, USA: Cambridge University Press, 2nd edition. ISBN 052189560X, 9780521895606.

Perdomo, J. C.; Britton, T.; Hardt, M.; and Abebe, R. 2023. Difficult Lessons on Social Prediction from Wisconsin Public Schools. *arXiv preprint arXiv:2304.06205*.

Peters, J.; Janzing, D.; and Schölkopf, B. 2017. *Elements of causal inference: foundations and learning algorithms*. The MIT Press.

Prosperi, M.; Guo, Y.; Sperrin, M.; Koopman, J. S.; Min, J. S.; He, X.; Rich, S.; Wang, M.; Buchan, I. E.; and Bian, J. 2020. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7): 369–375.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artificial intelligence*, 32(1): 57–95.

Rosenbaum, P. R.; Rosenbaum, P. B.; and Briskman. 2010. *Design of observational studies*, volume 10. Springer.

Rosenbaum, P. R.; and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1): 41–55.

Ross, A.; Lakkaraju, H.; and Bastani, O. 2021. Learning models for actionable recourse. *Advances in Neural Information Processing Systems*, 34: 18734–18746.

Rubin, D. B. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469): 322–331.

Saxena, D.; Moon, E. S.-Y.; Chaurasia, A.; Guan, Y.; and Guha, S. 2023. Rethinking "Risk" in Algorithmic Systems Through A Computational Narrative Analysis of Casenotes in Child-Welfare. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19.

Shalit, U.; Johansson, F. D.; and Sontag, D. 2017. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, 3076–3085. PMLR.

Shavit, Y.; Edelman, B.; and Axelrod, B. 2020. Causal strategic linear regression. In *International Conference on Machine Learning*, 8676–8686. PMLR.

Stephenson, J.; and Imrie, J. 1998. Why do we need randomised controlled trials to assess behavioural interventions? *BMJ*, 316(7131): 611–613.

Tamhane, A.; Ikbal, S.; Sengupta, B.; Duggirala, M.; and Appleton, J. 2014. Predicting student risks through longitudinal analysis. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1544–1552.

Ustun, B.; Spangher, A.; and Liu, Y. 2019. Actionable recourse in linear classification. In *Proceedings of the conference on fairness, accountability, and transparency*, 10–19.

Verma, S.; Boonsanong, V.; Hoang, M.; Hines, K. E.; Dickerson, J. P.; and Shah, C. 2020. Counterfactual explanations and algorithmic recourses for machine learning: A review. *arXiv preprint arXiv:2010.10596*.

Wager, S.; and Athey, S. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.

Wang, A.; Kapoor, S.; Barocas, S.; and Narayanan, A. 2022. Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy. *Available at SSRN*.

Xu, J.; Han, Y.; Marcu, D.; and Van Der Schaar, M. 2017. Progressive prediction of student performance in college programs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Yeager, D. S.; Hanselman, P.; Walton, G. M.; Murray, J. S.; Crosnoe, R.; Muller, C.; Tipton, E.; Schneider, B.; Hulleman, C. S.; Hinojosa, C. P.; et al. 2019. A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774): 364–369.

Zhao, Y.; Zeng, D.; Rush, A. J.; and Kosorok, M. R. 2012. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499): 1106–1118.