# Omniprediction

## One Predictor, Many Downstream Objectives

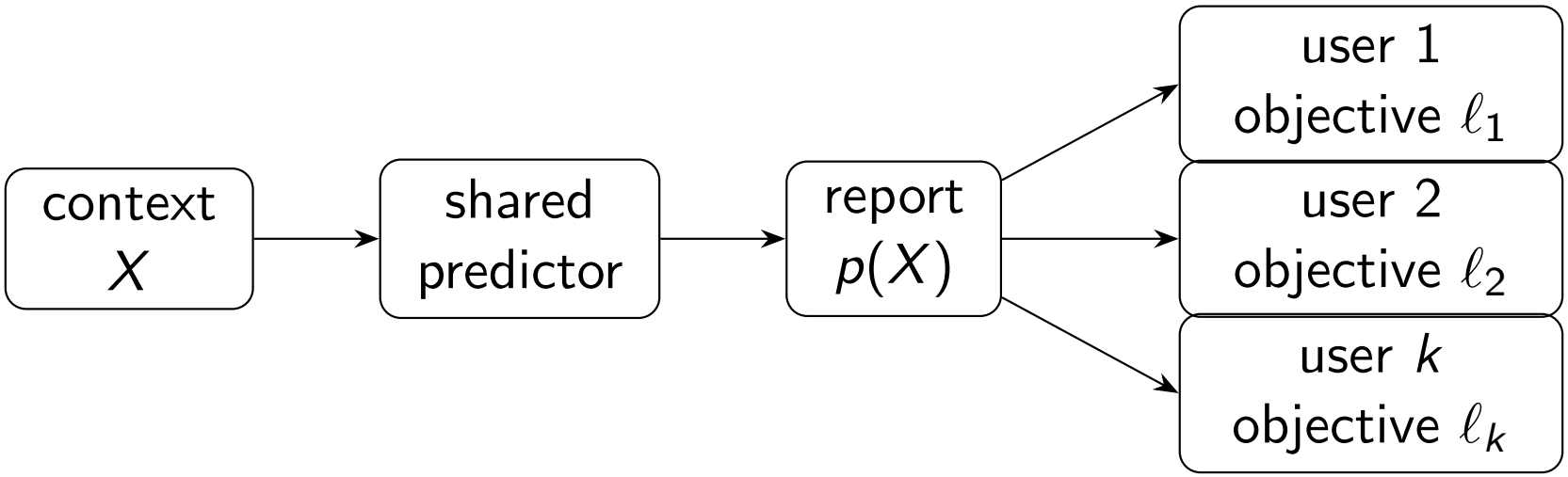
Aaron Roth

University of Pennsylvania

Fall 2026

# The Alignment Goal: One Predictor, Many Users

A single AI system observes a rich context $X$ and produces a probabilistic report $p(X)$. The same predictor is shared by decision makers with different objectives.




## Alignment as a decision interface

For every supported user, best responding to $p(X)$ should perform nearly as well as a benchmark policy that can use the original context $X$—with the same predictor serving every user's objective.

# Overview

- Calibration can guarantee low swap regret even when a forecast discards information that would improve decisions.

# Overview

- Calibration can guarantee low swap regret even when a forecast discards information that would improve decisions.
- Omniprediction asks one predictor to support many users while competing with policies that use the original context.

# Overview

- Calibration can guarantee low swap regret even when a forecast discards information that would improve decisions.
- Omniprediction asks one predictor to support many users while competing with policies that use the original context.
- Calibration plus context-sensitive checks gives this stronger guarantee, and approachability lets us construct such forecasts.

# Recall What Calibration Guarantees

Today the outcome is binary: $Y \in \{0, 1\}$. The context $X$ is available before prediction, and $p(X)$ reports the probability that $Y = 1$. We first work under a fixed population distribution.

# Recall What Calibration Guarantees

Today the outcome is binary: $Y \in \{0, 1\}$. The context $X$ is available before prediction, and $p(X)$ reports the probability that $Y = 1$. We first work under a fixed population distribution.
*Calibration:* prediction errors average to zero within each report cell:

$$\mathbb{E}[Y - p(X) \mid p(X)] = 0.$$

*Decision calibration:* for a fixed user's best-response policy $a_p$, errors average to zero within each action cell:

$$\mathbb{E}[Y - p(X) \mid a_p(X)] = 0.$$

# Recall What Calibration Guarantees

Today the outcome is binary: $Y \in \{0, 1\}$. The context $X$ is available before prediction, and $p(X)$ reports the probability that $Y = 1$. We first work under a fixed population distribution.
*Calibration:* prediction errors average to zero within each report cell:

$$\mathbb{E}[Y - p(X) \mid p(X)] = 0.$$

*Decision calibration:* for a fixed user's best-response policy $a_p$, errors average to zero within each action cell:

$$\mathbb{E}[Y - p(X) \mid a_p(X)] = 0.$$

Both group contexts using the predictor's own reports or the actions they induce. What if the predictor gives every context the same report?

# A Constant Mean Forecast Passes Both Tests

Suppose the context $X$ contains a diagnostic signal that perfectly reveals whether a proposal is safe, and the two states are equally likely:

| signal in $X$ | probability | outcome $Y$ | report $p(X)$ |
|---|---|---|---|
| indicates safe | $1/2$ | $1$ | $1/2$ |
| indicates harmful | $1/2$ | $0$ | $1/2$ |

# A Constant Mean Forecast Passes Both Tests

Suppose the context $X$ contains a diagnostic signal that perfectly reveals whether a proposal is safe, and the two states are equally likely:

| signal in $X$ | probability | outcome $Y$ | report $p(X)$ |
|---|---|---|---|
| indicates safe | $1/2$ | $1$ | $1/2$ |
| indicates harmful | $1/2$ | $0$ | $1/2$ |

The constant report is calibrated because

$$\mathbb{E}[Y \mid p(X) = 1/2] = \mathbb{E}[Y] = \frac{1}{2}.$$

# A Constant Mean Forecast Passes Both Tests

Suppose the context $X$ contains a diagnostic signal that perfectly reveals whether a proposal is safe, and the two states are equally likely:

| signal in $X$ | probability | outcome $Y$ | report $p(X)$ |
|---|---|---|---|
| indicates safe | $1/2$ | $1$ | $1/2$ |
| indicates harmful | $1/2$ | $0$ | $1/2$ |

The constant report is calibrated because

$$\mathbb{E}[Y \mid p(X) = 1/2] = \mathbb{E}[Y] = \frac{1}{2}.$$

It also induces the same action for both signal values. Its sole action cell contains the whole population, on which

$$\mathbb{E}[Y - p(X)] = 0.$$

Thus the forecast is decision calibrated for every downstream user.

# Low Swap Regret Can Be Weak

Consider this loss model. Deploying safely costs 0 and deploying harmfully costs 1. Holding incurs a modest delay or opportunity cost of $1/4$:

| | safe ($Y = 1$) | harmful ($Y = 0$) |
|---|---|---|
| deploy | 0 | 1 |
| hold | $1/4$ | $1/4$ |

Under the report $p(X) = 1/2$, the forecasted losses are

$$\text{deploy: } \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}, \qquad \text{hold: } \frac{1}{4},$$

so the best response policy always holds.

# Low Swap Regret Can Be Weak

Consider this loss model. Deploying safely costs 0 and deploying harmfully costs 1. Holding incurs a modest delay or opportunity cost of $1/4$:

| | safe ($Y = 1$) | harmful ($Y = 0$) |
|---|---|---|
| deploy | 0 | 1 |
| hold | $1/4$ | $1/4$ |

Under the report $p(X) = 1/2$, the forecasted losses are

$$\text{deploy: } \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}, \qquad \text{hold: } \frac{1}{4},$$

so the best response policy always holds.

The only change an action remapping can make on the realized hold cell is to replace every hold decision by deploy. Its true expected loss is $1/2$, rather than $1/4$. No remapping improves on hold, so swap regret is exactly zero.

# The Original Context Supports a Better Policy

A policy that uses the diagnostic signal can choose

$$c(X) = \begin{cases} \text{deploy}, & X \text{ indicates safe}, \\ \text{hold}, & X \text{ indicates harmful}. \end{cases}$$

Its expected loss is

$$\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} < \frac{1}{4},$$

better than the loss of always holding.

# The Original Context Supports a Better Policy

A policy that uses the diagnostic signal can choose

$$c(X) = \begin{cases} \text{deploy}, & X \text{ indicates safe}, \\ \text{hold}, & X \text{ indicates harmful}. \end{cases}$$

Its expected loss is

$$\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} < \frac{1}{4},$$

better than the loss of always holding.
There is no contradiction. Calibration and swap regret compare against rules that see only the report or the induced action. Those comparison rules cannot recover the distinction that the constant report discarded.

# Which Policies Should the Predictor Compete With?

A *benchmark class* is the set of alternative policies against which we measure a decision rule.

| benchmark class | information it sees | distinction it can make |
|---|---|---|
| action remappings $\phi(a_p(X))$ | induced action | only between action cells |
| contextual policies $c(X)$ | original context | between contexts receiving the same report |

# Which Policies Should the Predictor Compete With?

A *benchmark class* is the set of alternative policies against which we measure a decision rule.

| benchmark class | information it sees | distinction it can make |
|---|---|---|
| action remappings $\phi(a_p(X))$ | induced action | only between action cells |
| contextual policies $c(X)$ | original context | between contexts receiving the same report |

Contextual policies can give a more demanding comparison: they can use distinctions that the predictor ignored.

# One Predictor Serves Many Downstream Users

Let $(X, Y) \sim \mathcal{D}$, where $X \in \mathcal{X}$ and $Y \in \{0, 1\}$, and fix one predictor

$$p : \mathcal{X} \longrightarrow [0, 1].$$

A downstream user has an action set $\mathcal{A}$ and a bounded loss

$$\ell : \mathcal{A} \times \{0, 1\} \longrightarrow [0, 1].$$

# One Predictor Serves Many Downstream Users

Let $(X, Y) \sim \mathcal{D}$, where $X \in \mathcal{X}$ and $Y \in \{0, 1\}$, and fix one predictor

$$p : \mathcal{X} \longrightarrow [0, 1].$$

A downstream user has an action set $\mathcal{A}$ and a bounded loss

$$\ell : \mathcal{A} \times \{0, 1\} \longrightarrow [0, 1].$$

The predictor $p$ is common to all users. The user's objective affects only which action they choose from the report.

# Each Loss Defines Its Own Best Response

If the probability of $Y = 1$ were $q$, the expected loss of action $a$ would be

$$\ell(a; q) := (1 - q)\ell(a, 0) + q\ell(a, 1).$$

Fix a deterministic tie-breaking rule. The user's best response to forecast $q$ is

$$\mathrm{BR}_\ell(q) \in \arg\min_{a \in \mathcal{A}} \ell(a; q).$$

# Each Loss Defines Its Own Best Response

If the probability of $Y = 1$ were $q$, the expected loss of action $a$ would be

$$\ell(a; q) := (1 - q)\ell(a, 0) + q\ell(a, 1).$$

Fix a deterministic tie-breaking rule. The user's best response to forecast $q$ is

$$\mathrm{BR}_\ell(q) \in \arg\min_{a \in \mathcal{A}} \ell(a; q).$$

Treating the report $p(x)$ as the event probability gives the induced *best-response (or plug-in) policy*

$$\pi_{p,\ell}(x) := \mathrm{BR}_\ell(p(x)).$$

The same predictor can therefore be post-processed differently for different losses.

# Omniprediction Competes with Contextual Policies

Let $\mathcal{L}$ be a family of downstream losses and let $\mathcal{C}$ be a class of contextual policies $c : \mathcal{X} \to \mathcal{A}$.

# Omniprediction Competes with Contextual Policies

Let $\mathcal{L}$ be a family of downstream losses and let $\mathcal{C}$ be a class of contextual policies $c : \mathcal{X} \to \mathcal{A}$.

## Definition

The predictor $p$ is an $\varepsilon$-*omnipredictor* for $(\mathcal{L}, \mathcal{C})$ if, for every $\ell \in \mathcal{L}$,

$$\mathbb{E}[\ell(\pi_{p,\ell}(X), Y)] \leq \inf_{c \in \mathcal{C}} \mathbb{E}[\ell(c(X), Y)] + \varepsilon.$$

# Omniprediction Competes with Contextual Policies

Let $\mathcal{L}$ be a family of downstream losses and let $\mathcal{C}$ be a class of contextual policies $c : \mathcal{X} \to \mathcal{A}$.

## Definition

The predictor $p$ is an $\varepsilon$-*omnipredictor* for $(\mathcal{L}, \mathcal{C})$ if, for every $\ell \in \mathcal{L}$,

$$\mathbb{E}[\ell(\pi_{p,\ell}(X), Y)] \leq \inf_{c \in \mathcal{C}} \mathbb{E}[\ell(c(X), Y)] + \varepsilon.$$

The guarantee is simultaneous over the loss family:

| one fixed $p$, for every $\ell \in \mathcal{L}$ : $\pi_{p,\ell} = \mathrm{BR}_\ell \circ p$. |
|---|

# Imagine a World in Which the Predictor Is Correct

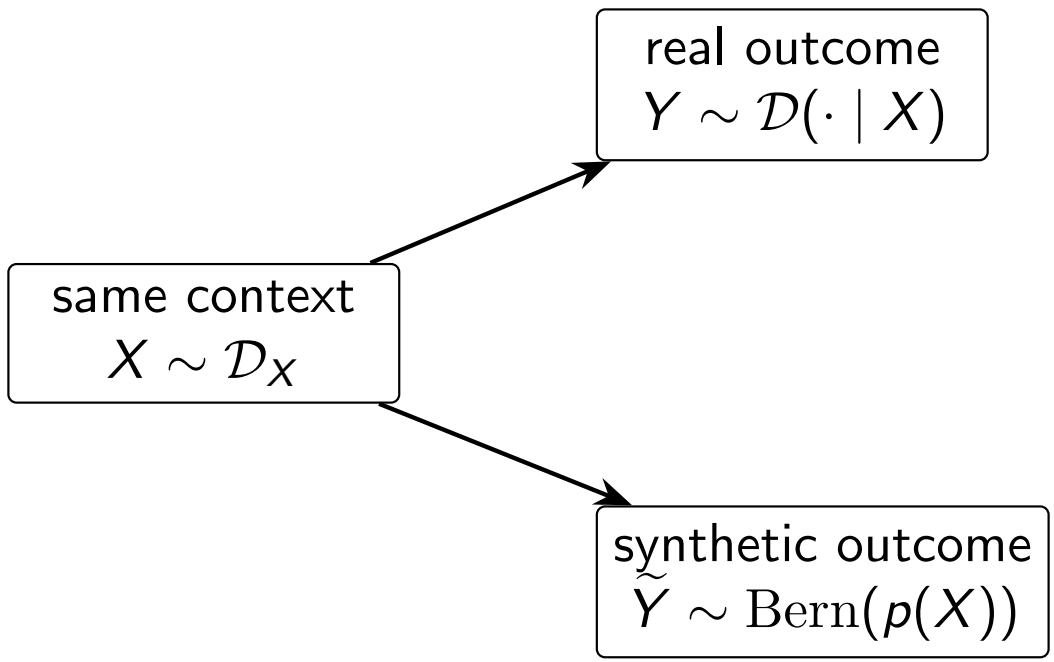
Keep the real distribution of contexts, but generate a synthetic outcome by

$$\widetilde{Y} \mid X = x \sim \text{Bernoulli}(p(x)).$$

# Imagine a World in Which the Predictor Is Correct

Keep the real distribution of contexts, but generate a synthetic outcome by

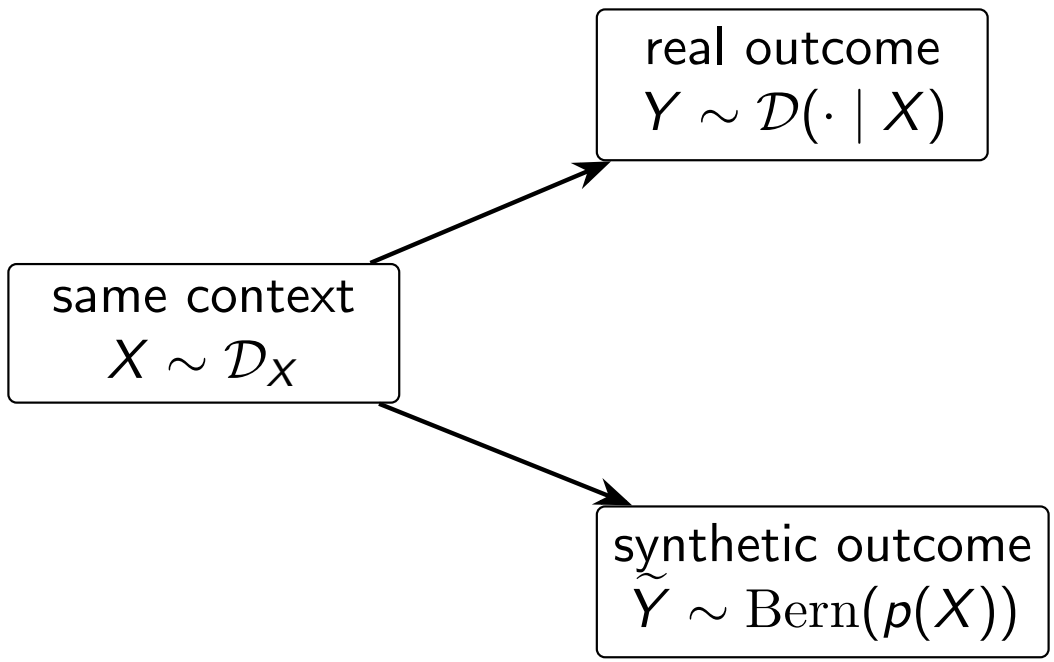
$$\widetilde{Y} \mid X = x \sim \text{Bernoulli}(p(x)).$$



# Imagine a World in Which the Predictor Is Correct

Keep the real distribution of contexts, but generate a synthetic outcome by

$$\widetilde{Y} \mid X = x \sim \text{Bernoulli}(p(x)).$$




The synthetic world is a proof device, not an assumption about the real data. In it, $p(x)$ is exactly the conditional probability of the outcome at every context $x$.

# Best Response Is Optimal in the Synthetic World

For a policy $\pi : \mathcal{X} \to \mathcal{A}$, define its real and synthetic risks by

$$R_\ell(\pi) := \mathbb{E}[\ell(\pi(X), Y)], \qquad \widetilde{R}_{\ell,\boldsymbol{p}}(\pi) := \mathbb{E}[\ell(\pi(X), \widetilde{Y})].$$

# Best Response Is Optimal in the Synthetic World

For a policy $\pi : \mathcal{X} \to \mathcal{A}$, define its real and synthetic risks by

$$R_\ell(\pi) := \mathbb{E}[\ell(\pi(X), Y)], \qquad \widetilde{R}_{\ell,p}(\pi) := \mathbb{E}[\ell(\pi(X), \widetilde{Y})].$$

At each context $x$, the synthetic probability of $Y = 1$ is $p(x)$. Therefore the definition of $\mathrm{BR}_\ell$ gives

$$\ell(\mathrm{BR}_\ell(p(x)); p(x)) \le \ell(\pi(x); p(x)) \qquad \text{for every policy } \pi.$$

Averaging over $X$ yields

$$\boxed{\widetilde{R}_{\ell,p}(\pi_{p,\ell}) \le \widetilde{R}_{\ell,p}(\pi) \quad \text{for every } \pi.}$$

# Two Real–Synthetic Comparisons Would Finish the Proof

Fix a loss $\ell$ and a contextual benchmark $c \in \mathcal{C}$. We want to justify this chain, with small errors $\alpha, \beta \geq 0$:

$$
\begin{aligned}
R_\ell(\pi_{\boldsymbol{p},\ell}) &\leq \widetilde{R}_{\ell,\boldsymbol{p}}(\pi_{\boldsymbol{p},\ell}) + \alpha && \text{(still to prove)} \\
&\leq \widetilde{R}_{\ell,\boldsymbol{p}}(c) + \alpha && \text{(already proved)} \\
&\leq R_\ell(c) + \alpha + \beta && \text{(still to prove).}
\end{aligned}
$$

# Two Real–Synthetic Comparisons Would Finish the Proof

Fix a loss $\ell$ and a contextual benchmark $c \in \mathcal{C}$. We want to justify this chain, with small errors $\alpha, \beta \geq 0$:

$$\begin{aligned} R_\ell(\pi_{\boldsymbol{p},\ell}) &\leq \widetilde{R}_{\ell,\boldsymbol{p}}(\pi_{\boldsymbol{p},\ell}) + \alpha && \text{(still to prove)} \\ &\leq \widetilde{R}_{\ell,\boldsymbol{p}}(c) + \alpha && \text{(already proved)} \\ &\leq R_\ell(c) + \alpha + \beta && \text{(still to prove).} \end{aligned}$$

The middle step is optimality in the synthetic world.
The two remaining tasks ask how much a policy's expected loss changes when we replace real outcomes by synthetic ones: first for the plug-in policy, then for the contextual benchmark.

# A Binary Loss Is Affine in the Outcome

Define the change in action $a$'s loss between the two outcomes:

$$\Delta_\ell(a) := \ell(a,1) - \ell(a,0).$$

Since $Y \in \{0,1\}$,

$$\begin{aligned}\ell(a,Y) &= (1-Y)\ell(a,0) + Y\ell(a,1)\\ &= \ell(a,0) + Y\big(\ell(a,1) - \ell(a,0)\big)\\ &= \ell(a,0) + Y\Delta_\ell(a).\end{aligned}$$

# A Binary Loss Is Affine in the Outcome

Define the change in action $a$'s loss between the two outcomes:

$$\Delta_\ell(a) := \ell(a,1) - \ell(a,0).$$

Since $Y \in \{0,1\}$,

$$\begin{aligned}\ell(a,Y) &= (1-Y)\ell(a,0) + Y\ell(a,1)\\ &= \ell(a,0) + Y\big(\ell(a,1) - \ell(a,0)\big)\\ &= \ell(a,0) + Y\Delta_\ell(a).\end{aligned}$$

In the synthetic world, conditional expectation replaces $Y$ by its synthetic mean $p(X)$:

$$\mathbb{E}[\ell(a,\widetilde{Y}) \mid X] = \ell(a,0) + p(X)\Delta_\ell(a).$$

# Every Transfer Error Is a Residual Correlation

Apply the preceding identities to the context-dependent action $\pi(X)$:

$$
\begin{aligned}
&R_\ell(\pi) - \widetilde{R}_{\ell,p}(\pi) \\
&\quad = \mathbb{E}\left[\ell(\pi(X), 0) + Y\Delta_\ell(\pi(X)) - \ell(\pi(X), 0) - p(X)\Delta_\ell(\pi(X))\right] \\
&\quad = \boxed{\mathbb{E}\left[(Y - p(X))\Delta_\ell(\pi(X))\right].}
\end{aligned}
$$

# Every Transfer Error Is a Residual Correlation

Apply the preceding identities to the context-dependent action $\pi(X)$:

$$
\begin{aligned}
&R_\ell(\pi) - \widetilde{R}_{\ell,p}(\pi) \\
&\quad = \mathbb{E}\left[\ell(\pi(X), 0) + Y\Delta_\ell(\pi(X)) - \ell(\pi(X), 0) - p(X)\Delta_\ell(\pi(X))\right] \\
&\quad = \boxed{\mathbb{E}\left[(Y - p(X))\Delta_\ell(\pi(X))\right].}
\end{aligned}
$$

The real and synthetic worlds agree on the loss of $\pi$ whenever the forecast residual $Y - p(X)$ has small correlation with the decision-dependent multiplier $\Delta_\ell(\pi(X))$.

# Calibration Transfers the Plug-In Policy

Define the population calibration error

$$\mathrm{Cal}(p) := \mathbb{E}\left[\left|\mathbb{E}[Y - p(X) \mid p(X)]\right|\right].$$

For the best response policy,

$$\Delta_\ell(\pi_{p,\ell}(X)) = \Delta_\ell(\mathrm{BR}_\ell(p(X)))$$

is a function only of the report $p(X)$, and its absolute value is at most one because $\ell \in [0, 1]$.

# Calibration Transfers the Plug-In Policy

Define the population calibration error

$$\mathrm{Cal}(p) := \mathbb{E}\left[\left|\mathbb{E}[Y - p(X) \mid p(X)]\right|\right].$$

For the best response policy,

$$\Delta_\ell(\pi_{p,\ell}(X)) = \Delta_\ell(\mathrm{BR}_\ell(p(X)))$$

is a function only of the report $p(X)$, and its absolute value is at most one because $\ell \in [0, 1]$. Conditioning first on $p(X)$ and using the tower property therefore gives

$$\begin{aligned}
&\left|R_\ell(\pi_{p,\ell}) - \widetilde{R}_{\ell,p}(\pi_{p,\ell})\right| \\
&= \left|\mathbb{E}\left[\mathbb{E}[Y - p(X) \mid p(X)]\Delta_\ell(\mathrm{BR}_\ell(p(X)))\right]\right| \\
&\leq \mathrm{Cal}(p).
\end{aligned}$$

Calibration controls the transfer of the best response policy betewen the two worlds.

# Calibration Does Not Transfer a Contextual Benchmark

For a contextual policy $c$, the relevant residual multiplier is

$$\Delta_\ell(c(X)).$$

It may depend on distinctions in $X$ that are absent from $p(X)$. We therefore cannot treat it as a function of the report when conditioning on $p(X)$.

# Calibration Does Not Transfer a Contextual Benchmark

For a contextual policy $c$, the relevant residual multiplier is

$$\Delta_\ell(c(X)).$$

It may depend on distinctions in $X$ that are absent from $p(X)$. We therefore cannot treat it as a function of the report when conditioning on $p(X)$.
This is exactly what happened in the constant-forecast example: the report had one calibrated cell, while the useful benchmark took different actions for the two signal values inside that cell.

# Calibration Does Not Transfer a Contextual Benchmark

For a contextual policy $c$, the relevant residual multiplier is

$$\Delta_\ell(c(X)).$$

It may depend on distinctions in $X$ that are absent from $p(X)$. We therefore cannot treat it as a function of the report when conditioning on $p(X)$.
This is exactly what happened in the constant-forecast example: the report had one calibrated cell, while the useful benchmark took different actions for the two signal values inside that cell.
To compete with contextual policies, we need residual tests that retain those context-dependent distinctions.

# A Contextual Test Detects the Missing Information

Recall that deploy loses $1 - Y$, while hold always loses $1/4$. Thus

$$\Delta_\ell(\text{deploy}) = -1, \qquad \Delta_\ell(\text{hold}) = 0.$$

For any deploy-or-hold benchmark $c$, the loss comparison uses the test

$$w_c(X) = \Delta_\ell(c(X)) = -\mathbf{1}\{c(X) = \text{deploy}\}.$$

We need prediction errors to cancel where *this benchmark deploys*, not just where the predictor gives the same report.

# A Contextual Test Detects the Missing Information

Recall that deploy loses $1 - Y$, while hold always loses $1/4$. Thus

$$\Delta_\ell(\text{deploy}) = -1, \qquad \Delta_\ell(\text{hold}) = 0.$$

For any deploy-or-hold benchmark $c$, the loss comparison uses the test

$$w_c(X) = \Delta_\ell(c(X)) = -\mathbf{1}\{c(X) = \text{deploy}\}.$$

We need prediction errors to cancel where *this benchmark deploys*, not just where the predictor gives the same report.
Our benchmark deploys in the safe half of contexts and holds in the harmful half. For the constant forecast $p(X) = 1/2$,

$$\begin{aligned}\mathbb{E}[(Y - p(X))w_c(X)] &= \frac{1}{2}\left(1 - \frac{1}{2}\right)(-1) + \frac{1}{2}\left(0 - \frac{1}{2}\right)(0) \\ &= -\frac{1}{4}.\end{aligned}$$

This contextual test detects the distinction calibration missed.

# Multiaccuracy Supplies the Contextual Tests

To generalize this requirement, fix a test family $\mathcal{W}$ of functions $w : \mathcal{X} \to [-1, 1]$ and define

$$\mathrm{MA}_{\mathcal{W}}(p) := \sup_{w \in \mathcal{W}} \left| \mathbb{E}[(Y - p(X))w(X)] \right| .$$

Small multiaccuracy means that the forecast residual has small correlation with every selected context-dependent test.

# Multiaccuracy Supplies the Contextual Tests

To generalize this requirement, fix a test family $\mathcal{W}$ of functions $w : \mathcal{X} \to [-1, 1]$ and define

$$\mathrm{MA}_{\mathcal{W}}(p) := \sup_{w \in \mathcal{W}} \left| \mathbb{E}[(Y - p(X))w(X)] \right| .$$

Small multiaccuracy means that the forecast residual has small correlation with every selected context-dependent test.
Each loss–benchmark pair supplies one test, just as deploy-or-hold did:

$$\mathcal{W}(\mathcal{L}, \mathcal{C}) := \{x \longmapsto \Delta_\ell(c(x)) : \ell \in \mathcal{L},\ c \in \mathcal{C}\} .$$

The residual identity then gives, for every $\ell \in \mathcal{L}$ and $c \in \mathcal{C}$,

$$\left| R_\ell(c) - \widetilde{R}_{\ell,p}(c) \right| \leq \mathrm{MA}_{\mathcal{W}(\mathcal{L},\mathcal{C})}(p).$$

# Calibration Plus Multiaccuracy Gives Omniprediction

**Theorem**

*If*

$$\mathsf{Cal}(p) \leq \alpha \qquad \textit{and} \qquad \mathsf{MA}_{\mathcal{W}(\mathcal{L},\mathcal{C})}(p) \leq \beta,$$

*then $p$ is an $(\alpha+\beta)$-omnipredictor for $(\mathcal{L},\mathcal{C})$.*

# Calibration Plus Multiaccuracy Gives Omniprediction

**Theorem**

*If*

$$\mathsf{Cal}(p) \leq \alpha \qquad \textit{and} \qquad \mathsf{MA}_{\mathcal{W}(\mathcal{L},\mathcal{C})}(p) \leq \beta,$$

*then $p$ is an $(\alpha+\beta)$-omnipredictor for $(\mathcal{L},\mathcal{C})$.*

The two conditions justify our two real–synthetic loss comparisons:

- calibration controls the change in the best response policies' expected losses;
- multiaccuracy controls the change in the contextual benchmarks' expected losses.

# Proof: Transfer Optimality Through the Synthetic World

Fix $\ell \in \mathcal{L}$ and $c \in \mathcal{C}$. The two transfer bounds and synthetic optimality give

$$\begin{aligned} R_\ell(\pi_{\boldsymbol{p},\ell}) &\leq \widetilde{R}_{\ell,\boldsymbol{p}}(\pi_{\boldsymbol{p},\ell}) + \alpha && \text{(calibration)} \\ &\leq \widetilde{R}_{\ell,\boldsymbol{p}}(c) + \alpha && \text{(synthetic optimality)} \\ &\leq R_\ell(c) + \alpha + \beta && \text{(multiaccuracy).} \end{aligned}$$

# Proof: Transfer Optimality Through the Synthetic World

Fix $\ell \in \mathcal{L}$ and $c \in \mathcal{C}$. The two transfer bounds and synthetic optimality give

$$\begin{aligned} R_\ell(\pi_{\boldsymbol{p},\ell}) &\leq \widetilde{R}_{\ell,\boldsymbol{p}}(\pi_{\boldsymbol{p},\ell}) + \alpha && \text{(calibration)} \\ &\leq \widetilde{R}_{\ell,\boldsymbol{p}}(c) + \alpha && \text{(synthetic optimality)} \\ &\leq R_\ell(c) + \alpha + \beta && \text{(multiaccuracy).} \end{aligned}$$

Because this holds for every $c \in \mathcal{C}$, take the infimum over $c$. Because $\ell \in \mathcal{L}$ was arbitrary, the same fixed predictor serves every loss in the family. $\square$

# Outcome Indistinguishability Names the Proof Principle

The proof did not require the real and synthetic distributions to be close in every respect. It required them to give nearly the same expected losses:

$$\left|\mathbb{E}[\ell(\pi(X), Y)] - \mathbb{E}[\ell(\pi(X), \widetilde{Y})]\right|$$

for the best response policy as well as for the benchmark policies.

# Outcome Indistinguishability Names the Proof Principle

The proof did not require the real and synthetic distributions to be close in every respect. It required them to give nearly the same expected losses:

$$\left|\mathbb{E}[\ell(\pi(X), Y)] - \mathbb{E}[\ell(\pi(X), \widetilde{Y})]\right|$$

for the best response policy as well as for the benchmark policies. This is *outcome indistinguishability* for these loss tests: replacing the real outcome by a draw from the forecast barely changes the evaluations that matter to our decision guarantee.
Calibration and multiaccuracy are sufficient ways to enforce these comparisons. For a finite collection of users and benchmarks, we can also enforce the needed comparisons directly using Lecture 8.

# Construct Forecasts That Pass the Needed Tests

Now return to Lecture 8's sequential setting, with *finite* loss and benchmark families $\mathcal{L}, \mathcal{C}$. On each round:

1. Observe the context $x_t$ and announce a distribution over forecasts.
2. The environment commits to $y_t$ before the forecast $p_t$ is drawn.
3. Reveal $p_t$ to users; then observe $y_t$ and update.

# Construct Forecasts That Pass the Needed Tests

Now return to Lecture 8's sequential setting, with *finite* loss and benchmark families $\mathcal{L}, \mathcal{C}$. On each round:

1. Observe the context $x_t$ and announce a distribution over forecasts.
2. The environment commits to $y_t$ before the forecast $p_t$ is drawn.
3. Reveal $p_t$ to users; then observe $y_t$ and update.

The loss comparisons require exactly these residual multipliers:

$$h_\ell^{\text{plug}}(x, p) := \Delta_\ell(\text{BR}_\ell(p)),$$
$$h_{\ell,c}^{\text{bench}}(x, p) := \Delta_\ell(c(x)).$$

List them as $h_1, \ldots, h_m$, all bounded by one in absolute value. We want the average of $(y_t - p_t)h_j(x_t, p_t)$ to be small for every test.

# The Residual Constraints Are Response Satisfiable

Use a finite grid with a forecast within $\eta > 0$ of every $q \in [0, 1]$. For the observed context $x$, define

$$g_j(x, p, y) := (y - p)h_j(x, p), \qquad C_\eta := [-\eta, \eta]^m.$$

For any hypothetical outcome mean $q$, choose a grid forecast $p$ with $|p - q| \leq \eta$. Then every coordinate satisfies

$$\left|\mathbb{E}_{Y \sim \mathrm{Bern}(q)}[g_j(x, p, Y)]\right| = |q - p|\,|h_j(x, p)| \leq \eta.$$

# The Residual Constraints Are Response Satisfiable

Use a finite grid with a forecast within $\eta > 0$ of every $q \in [0, 1]$. For the observed context $x$, define

$$g_j(x, p, y) := (y - p)h_j(x, p), \qquad C_\eta := [-\eta, \eta]^m.$$

For any hypothetical outcome mean $q$, choose a grid forecast $p$ with $|p - q| \leq \eta$. Then every coordinate satisfies

$$\left|\mathbb{E}_{Y \sim \mathrm{Bern}(q)}[g_j(x, p, Y)]\right| = |q - p|\,|h_j(x, p)| \leq \eta.$$

Thus $C_\eta$ is response satisfiable at every context. Lecture 8's minimax step applies after observing $x_t$, with the same Euclidean distance bound. Writing $\bar{g}_T = T^{-1}\sum_t g(x_t, p_t, y_t)$ gives

$$e_T := \max_j |\bar{g}_{T,j}| \leq \eta + \mathrm{dist}_2(\bar{g}_T, C_\eta).$$

Consequently $\mathbb{E}[e_T] \leq \eta + \varepsilon_T$, where $\varepsilon_T \to 0$ as $T \to \infty$. The expectation is over the forecaster's randomization.

# One Forecast Sequence Serves Every User

For a fixed loss $\ell$, write $a_t^\ell := \mathrm{BR}_\ell(p_t)$. Using the forecasted loss $\ell(a; p)$, the same three-step proof compares time averages:

$$\begin{aligned}\frac{1}{T}\sum_t \ell(a_t^\ell, y_t) &\le \frac{1}{T}\sum_t \ell(a_t^\ell; p_t) + e_T \\ &\le \frac{1}{T}\sum_t \ell(c(x_t); p_t) + e_T \\ &\le \frac{1}{T}\sum_t \ell(c(x_t), y_t) + 2e_T.\end{aligned}$$

The first and last steps use the plug-in and benchmark residual tests. The middle step is best responding to each report.

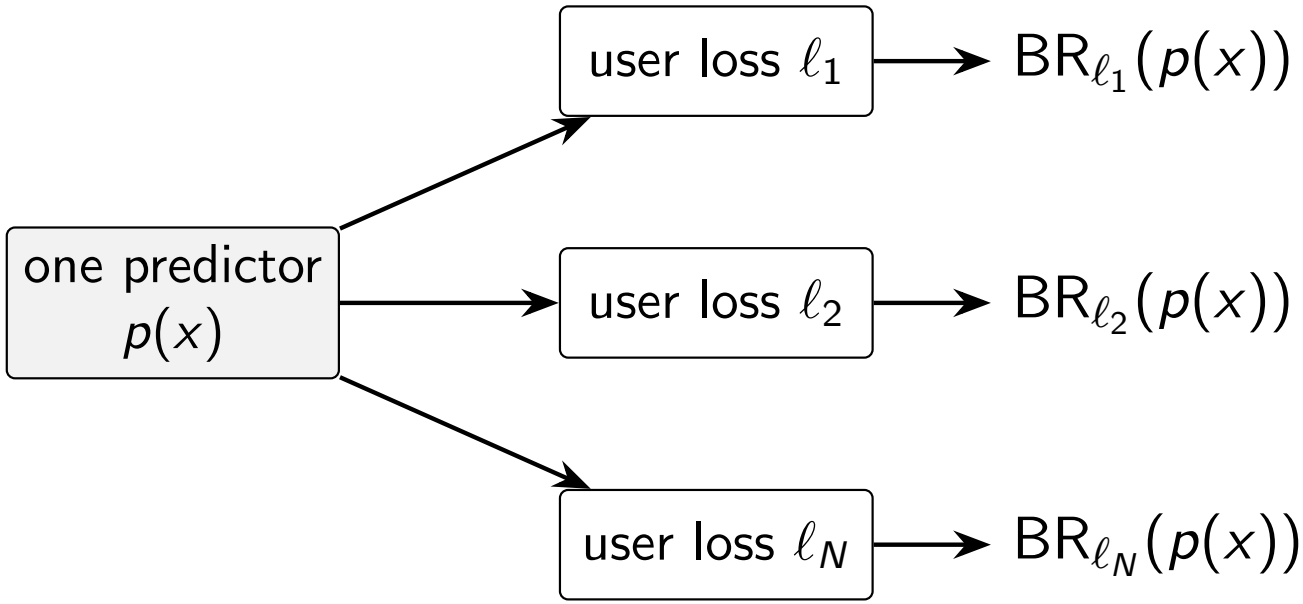
# One Forecast Sequence Serves Every User

For a fixed loss $\ell$, write $a_t^\ell := \mathrm{BR}_\ell(p_t)$. Using the forecasted loss $\ell(a; p)$, the same three-step proof compares time averages:

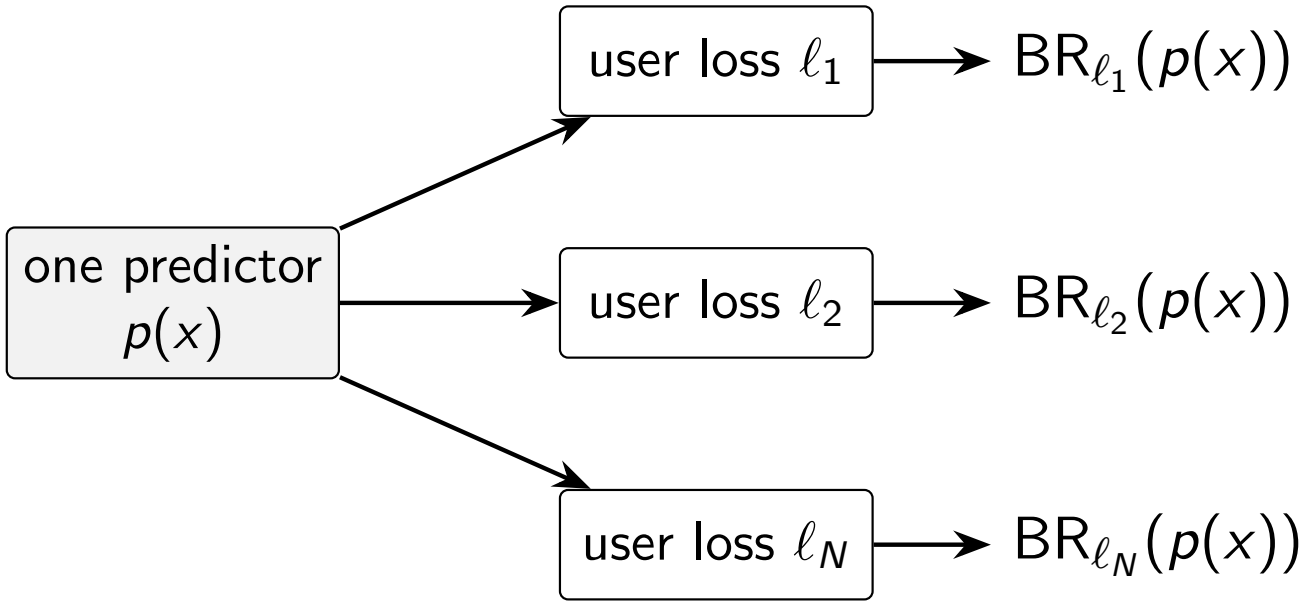
$$\begin{aligned}\frac{1}{T}\sum_t \ell(a_t^\ell, y_t) &\le \frac{1}{T}\sum_t \ell(a_t^\ell; p_t) + e_T \\ &\le \frac{1}{T}\sum_t \ell(c(x_t); p_t) + e_T \\ &\le \frac{1}{T}\sum_t \ell(c(x_t), y_t) + 2e_T.\end{aligned}$$

The first and last steps use the plug-in and benchmark residual tests. The middle step is best responding to each report.
This holds for every $\ell \in \mathcal{L}$ and $c \in \mathcal{C}$ on the realized sequence.
Since $\mathbb{E}[e_T] \le \eta + \varepsilon_T$, the expected excess loss is at most $2(\eta + \varepsilon_T)$, simultaneously for all supported users and benchmarks.
All users act on the *same* reports $p_1, \ldots, p_T$. The guarantee is for this sequence, not a fixed population predictor.

# One Predictor Becomes a Reusable Decision Interface



# One Predictor Becomes a Reusable Decision Interface




The alignment payoff is objective flexibility: different users can convert one report according to different objectives while retaining a contextual performance guarantee.

# Summary

- Calibration can miss useful information that a forecast discards.

# Summary

- Calibration can miss useful information that a forecast discards.
- Calibration plus context-sensitive checks lets one predictor support many objectives, with decisions competitive with contextual dependent benchmarks.

# Summary

- Calibration can miss useful information that a forecast discards.
- Calibration plus context-sensitive checks lets one predictor support many objectives, with decisions competitive with contextual dependent benchmarks.
- Approachability makes these guarantees constructive: one sequence of forecasts can serve many users even in Lecture 8's adversarial setting.

# Thanks!

See you next class!

# References


- Parikshit Gopalan, Adam Tauman Kalai, Omer Reingold, Vatsal Sharan, and Udi Wieder. "Omnipredictors." *ITCS*, 2022.
- Parikshit Gopalan, Lunjia Hu, Michael P. Kim, Omer Reingold, and Udi Wieder. "Loss Minimization through the Lens of Outcome Indistinguishability." *ITCS*, 2023.
- Michael P. Kim, Amirata Ghorbani, and James Zou. "Multiaccuracy: Black-Box Post-Processing for Fairness in Classification." *AAAI/ACM AIES*, 2019.
- David Blackwell. "An Analog of the Minimax Theorem for Vector Payoffs." *Pacific Journal of Mathematics*, 1956.