Panprediction

Optimal Predictions for Any Downstream Task and Loss

Siva Balakrishnan (Carnegie Mellon)
Nika Haghtalab (Berkeley)
Daniel Hsu (Columbia)
Brian Lee (Berkeley)
Eric Zhao (Berkeley)

Optimal Predictions for Any Downstream Task and Loss

What does this mean?

Optimal Predictions for Any Downstream Task and Loss

What does this mean?

Optimal Predictions for Any Downstream Task and Loss

What does this mean?

What does this mean?

Optimal Predictions for Any Downstream Task and Loss

What does this mean? What does this mean?

What does this mean?

Optimal Predictions for Any Downstream Task and Loss

What does this mean? **What does this mean?**

Isn’t this a calibration session?

Task (and Loss)

Fix:

- Features $X \in \mathcal{X}$
- Binary labels $Y \in \{0,1\}$
- Joint distribution D

Task (and Loss)

Fix:

- **Patient health records** $X \in \mathcal{X}$
- Binary labels $Y \in \{0,1\}$
- Joint distribution D

Task (and Loss)

Fix:

- **Patient health records** $X \in \mathcal{X}$
- **Cardiac event in 24 hours** $Y \in \{0,1\}$
- Joint distribution D

Task (and Loss)

Fix:

- **Patient health records** $X \in \mathcal{X}$
- **Cardiac event in 24 hours** $Y \in \{0,1\}$
- Joint distribution D

Predict **probability** p of **cardiac event** using **health records**

Loss $\ell \in \mathcal{L}$ compares p vs Y

Task (and Loss)

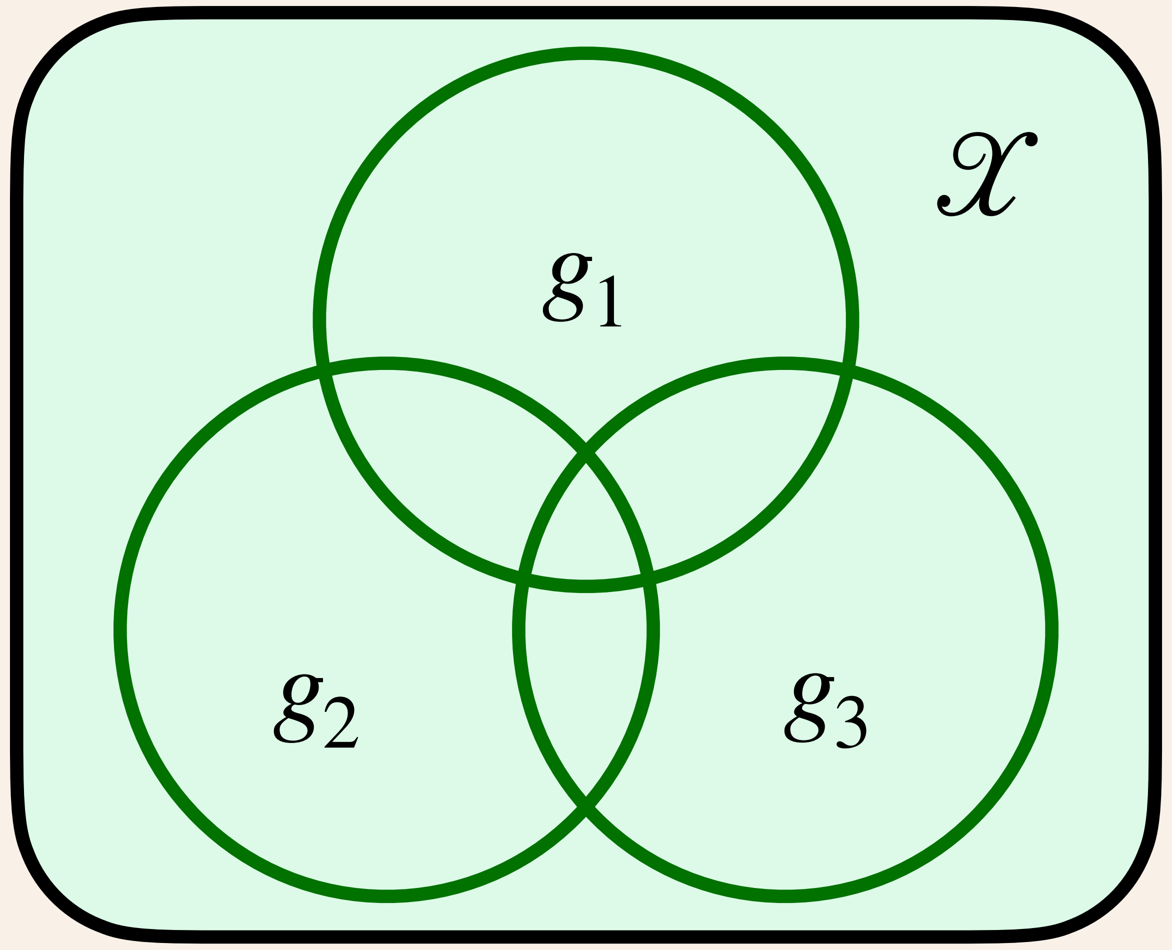
Fix:

- **Patient health records** $X \in \mathcal{X}$
- **Cardiac event in 24 hours** $Y \in \{0,1\}$
- Joint distribution D

Predict **probability** p of **cardiac event** using **health records**

Loss $\ell \in \mathscr{L}$ compares p vs Y

Groups $\mathcal{G} \subseteq 2^{\mathcal{X}}$



Task (and Loss)

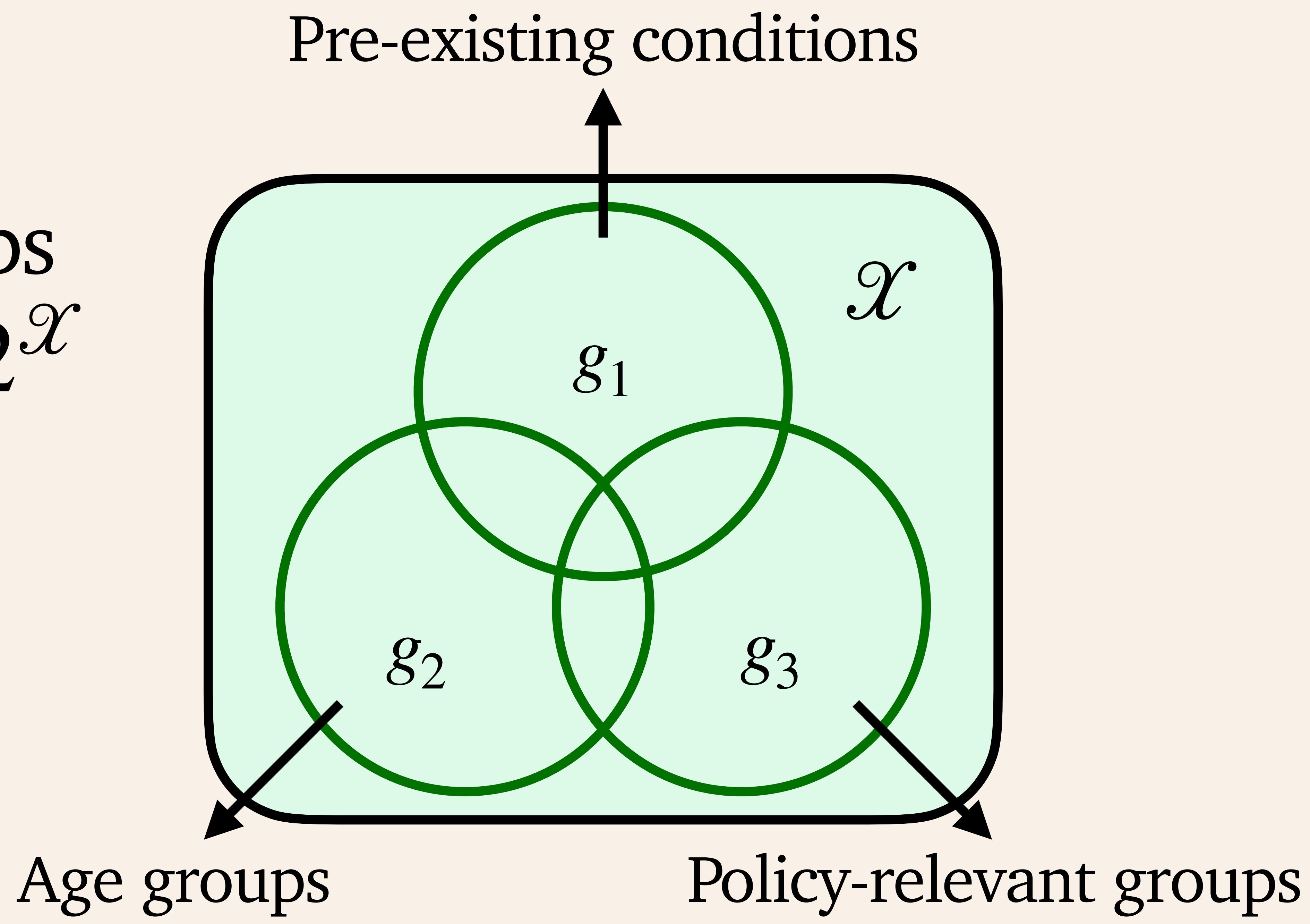
Fix:

- **Patient health records** $X \in \mathscr{X}$
- **Cardiac event in 24 hours** $Y \in \{0,1\}$
- Joint distribution D

Predict **probability** p of **cardiac event** using **health records**

Loss $\ell \in \mathscr{L}$ compares p vs Y

Groups $\mathscr{G} \subseteq 2^{\mathscr{X}}$



Task (and Loss)

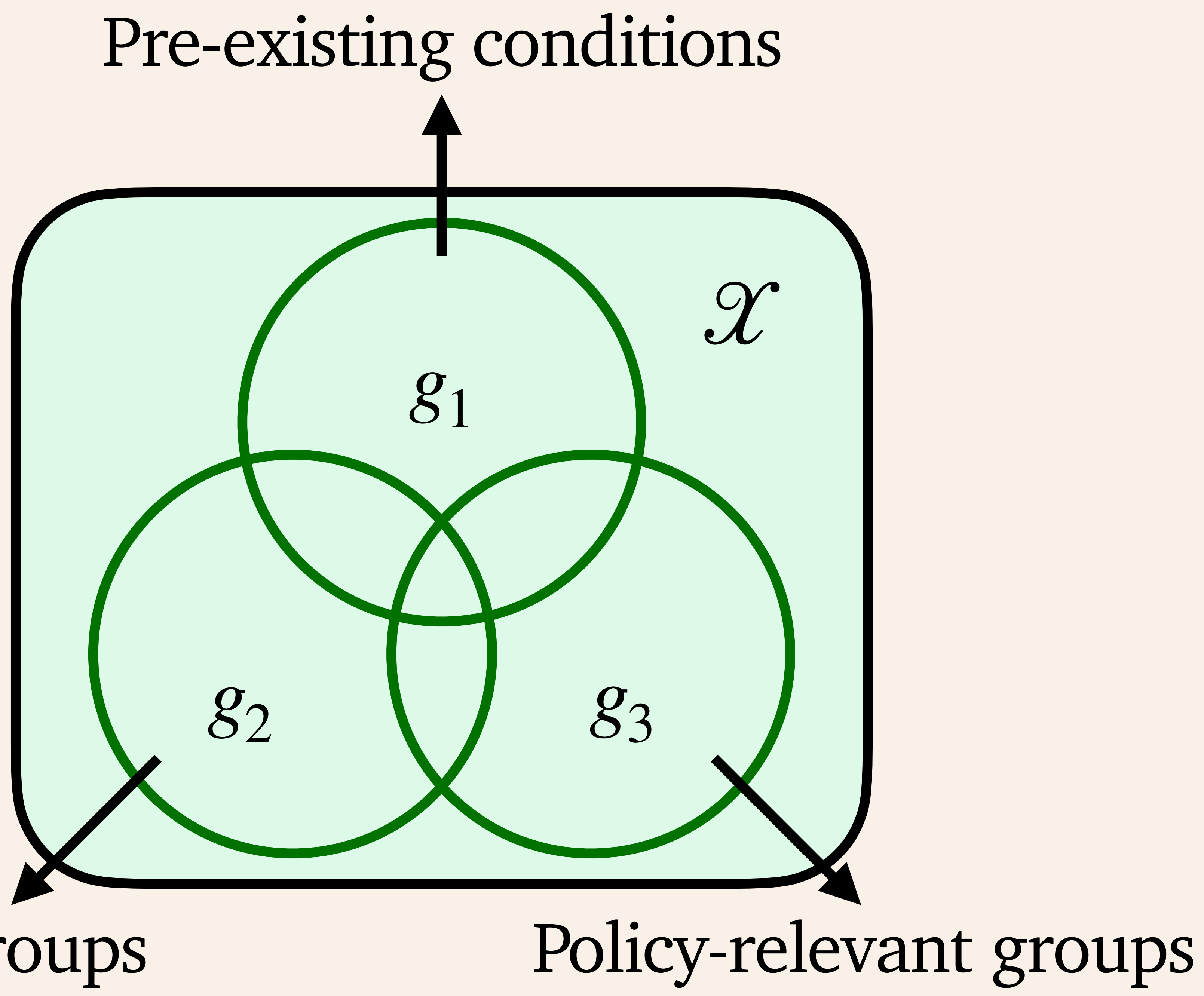
Fix:

- **Patient health records** $X \in \mathcal{X}$
- **Cardiac event in 24 hours** $Y \in \{0,1\}$
- Joint distribution D

Predict **probability** p of **cardiac event** using **health records**

Loss $\ell \in \mathcal{L}$ compares p vs Y

Groups $\mathcal{G} \subseteq 2^{\mathcal{X}}$



If X is in group $g \in \mathcal{G}$, say $g(X) = 1$

A **task** induced by g is the group-conditional distribution $D_g = D\,|_{g(X)=1}$

(Multi-group learning [RY21][TH22])

Optimality

Competitor hypothesis class $\mathscr{H} \subseteq \{h : \mathscr{X} \to [0,1]\}$

Optimality

Competitor hypothesis class $\mathscr{H} \subseteq \{h : \mathscr{X} \to [0,1]\}$

Physician

(ℓ_1, g_1)

Discharge coordinator

(ℓ_2, g_2)

Actuary

(ℓ_3, g_3)

Optimality

Competitor hypothesis class $\mathscr{H} \subseteq \{h : \mathscr{X} \to [0,1]\}$

Physician (ℓ_1, g_1)	→ $h_1 = \arg\min_{h \in H} \mathbb{E}_{(X,Y)\sim D_{g_1}} \left[\ell_1(h(X), Y)\right]$
Discharge coordinator (ℓ_2, g_2)	→ $h_2 = \arg\min_{h \in H} \mathbb{E}_{(X,Y)\sim D_{g_2}} \left[\ell_2(h(X), Y)\right]$
Actuary (ℓ_3, g_3)	→ $h_3 = \arg\min_{h \in H} \mathbb{E}_{(X,Y)\sim D_{g_3}} \left[\ell_3(h(X), Y)\right]$

Optimality

Competitor hypothesis class $\mathscr{H} \subseteq \{h : \mathscr{X} \to [0,1]\}$

Physician (ℓ_1, g_1) → $h_1 = \arg\min_{h \in H} \mathbb{E}_{(X,Y)\sim D_{g_1}} \left[\ell_1(h(X), Y)\right]$

Discharge coordinator (ℓ_2, g_2) → $h_2 = \arg\min_{h \in H} \mathbb{E}_{(X,Y)\sim D_{g_2}} \left[\ell_2(h(X), Y)\right]$

Actuary (ℓ_3, g_3) → $h_3 = \arg\min_{h \in H} \mathbb{E}_{(X,Y)\sim D_{g_3}} \left[\ell_3(h(X), Y)\right]$

Want our model to compete with h_1, h_2, h_3 **simultaneously**

Optimality

Competitor hypothesis class $\mathscr{H} \subseteq \{h : \mathscr{X} \to [0,1]\}$

Physician

(ℓ_1, g_1)

Discharge coordinator

(ℓ_2, g_2)

Actuary

(ℓ_3, g_3)

Say our model $\hat{h}$ is ε-**optimal** with respect to $\mathscr{H}$ if for every group $g \in \mathscr{G} = \{g_1, g_2, g_3\}$ and loss $\ell \in \mathscr{L} = \{\ell_1, \ell_2, \ell_3\}$,

$$\mathop{\mathbb{E}}_{(X,Y)\sim D_g}\left[\ell(\hat{h}(X), Y)\right] \approx_\varepsilon \min_{h\in H} \mathop{\mathbb{E}}_{(X,Y)\sim D_g}\left[\ell(h(X), Y)\right].$$

Optimality

Competitor hypothesis class $\mathcal{H} \subseteq \{h : \mathcal{X} \to [0,1]\}$

Physician

(ℓ_1, g_1)

Discharge coordinator

(ℓ_2, g_2)

Actuary

(ℓ_3, g_3)

Say our model $\hat{h}$ is **ε-optimal** with respect to $\mathcal{H}$ if for every group $g \in \mathcal{G} = \{g_1, g_2, g_3\}$ and loss $\ell \in \mathcal{L} = \{\ell_1, \ell_2, \ell_3\}$,

$$\mathop{\mathbb{E}}_{(X,Y)\sim D_g} \left[\ell(\hat{h}(X), Y)\right] \approx_\varepsilon \min_{h \in H} \mathop{\mathbb{E}}_{(X,Y)\sim D_g} \left[\ell(h(X), Y)\right].$$

We only know $(\mathcal{L}, \mathcal{G}, \mathcal{H})$

Competitor adapts to ℓ and g

Upstream vs Downstream

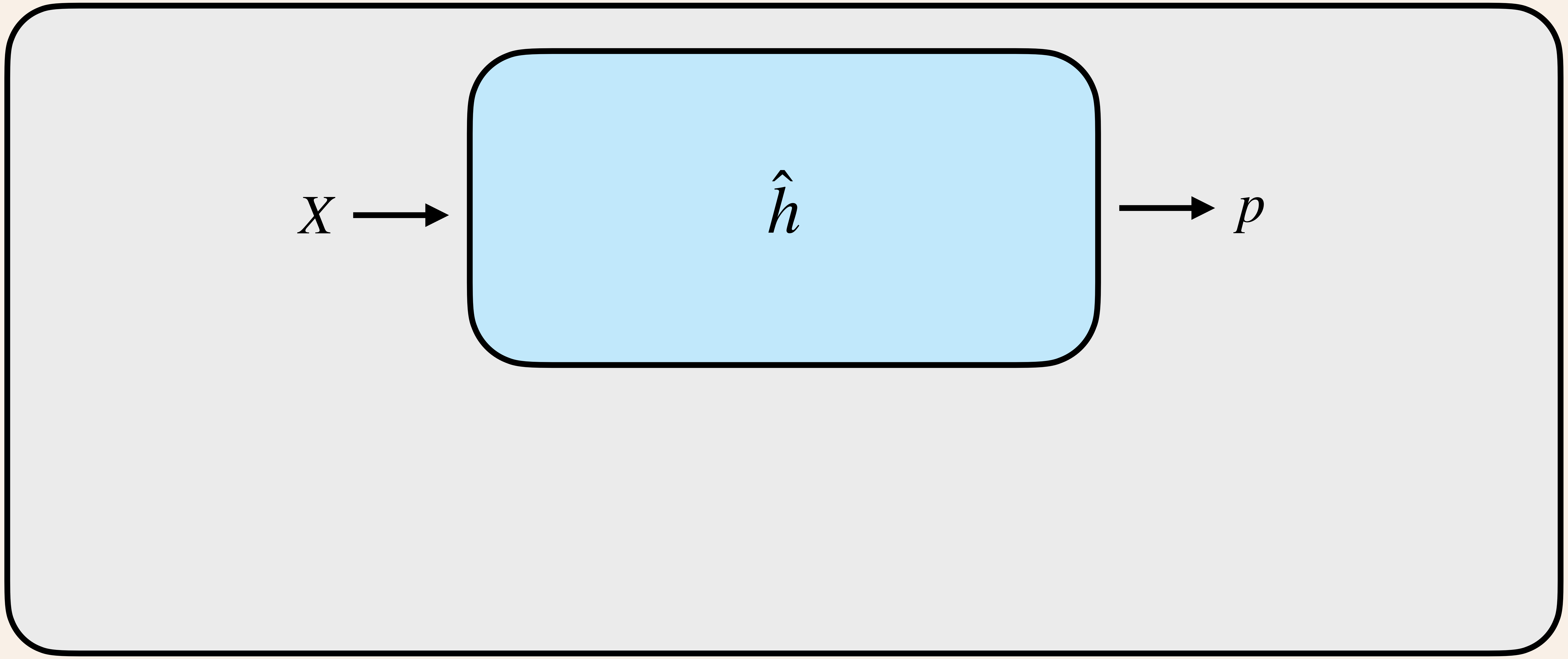
How construct magic model $\hat{h}$, if $h_1 \neq h_2 \neq h_3$?

Upstream vs Downstream

How construct magic model $\hat{h}$, if $h_1 \neq h_2 \neq h_3$? **Cheat (a bit).**

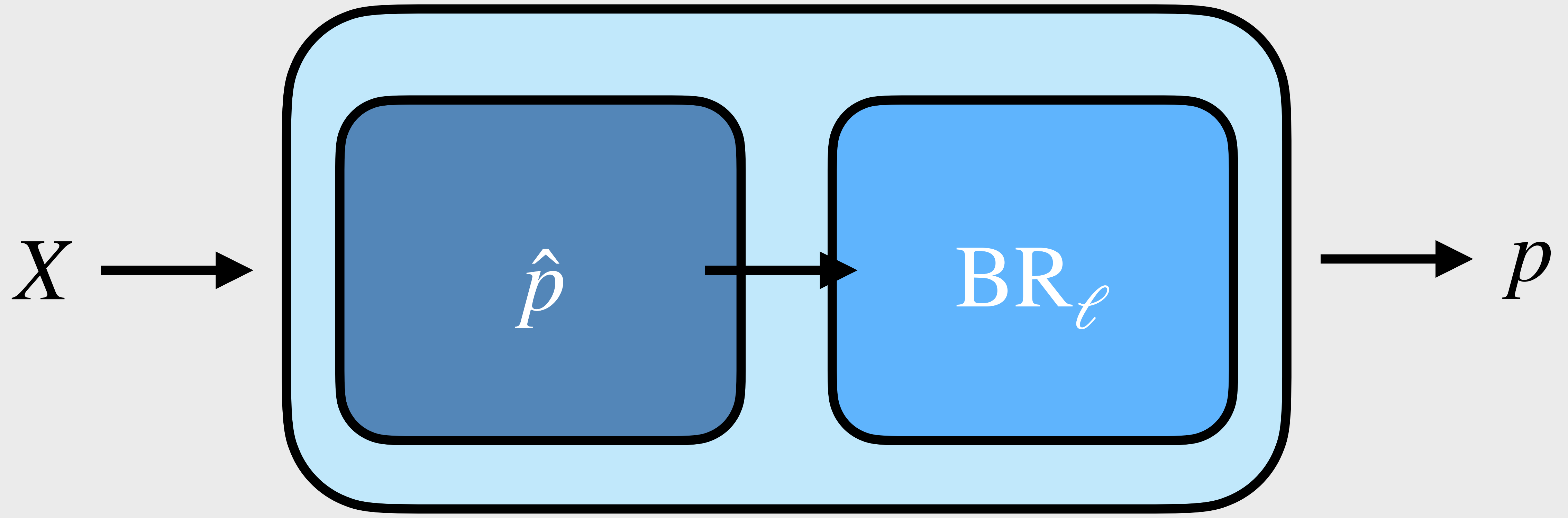
Upstream vs Downstream

How construct magic model $\hat{h}$, if $h_1 \neq h_2 \neq h_3$? **Cheat (a bit).**



Upstream vs Downstream

How construct magic model $\hat{h}$, if $h_1 \neq h_2 \neq h_3$? **Cheat (a bit).**



Upstream

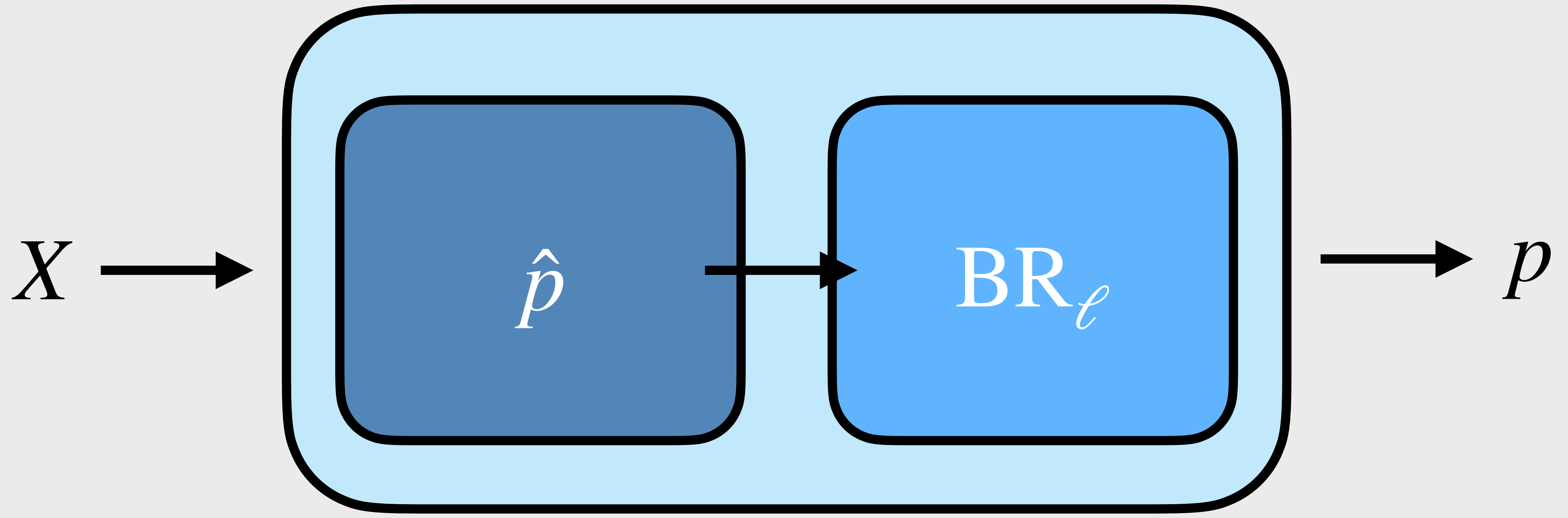
$$\hat{p} : \mathcal{X} \to [0,1]$$

Downstream

$$\mathrm{BR}_\ell(\hat{p}(X)) = \arg\min_{q \in [0,1]} \mathop{\mathbb{E}}_{W \sim \mathrm{Ber}(\hat{p}(X))} \left[\ell(q, W)\right]$$

Upstream vs Downstream

How construct magic model $\hat{h}$, if $h_1 \neq h_2 \neq h_3$? **Cheat (a bit).**



Upstream

$$\hat{p} : \mathcal{X} \to [0,1]$$

Downstream

$$\mathrm{BR}_\ell(\hat{p}(X)) = \arg\min_{q \in [0,1]} \mathop{\mathbb{E}}_{W \sim \mathrm{Ber}(\hat{p}(X))} \left[\ell(q, W)\right]$$

For each loss $\ell \in \mathscr{L}$, we have $\hat{h}_\ell = \mathrm{BR}_\ell \circ \hat{p}$. But only $\hat{p}$ needs to be **learned**.

(Omniprediction [GKR+22][GHK+22][OKK25])

Panprediction

Fix $(\mathscr{L}, \mathscr{G}, \mathscr{H})$.

Learn a predictor $\hat{p} : \mathscr{X} \to [0,1]$ such that for all $g \in \mathscr{G}$ and $\ell \in \mathscr{L}$,

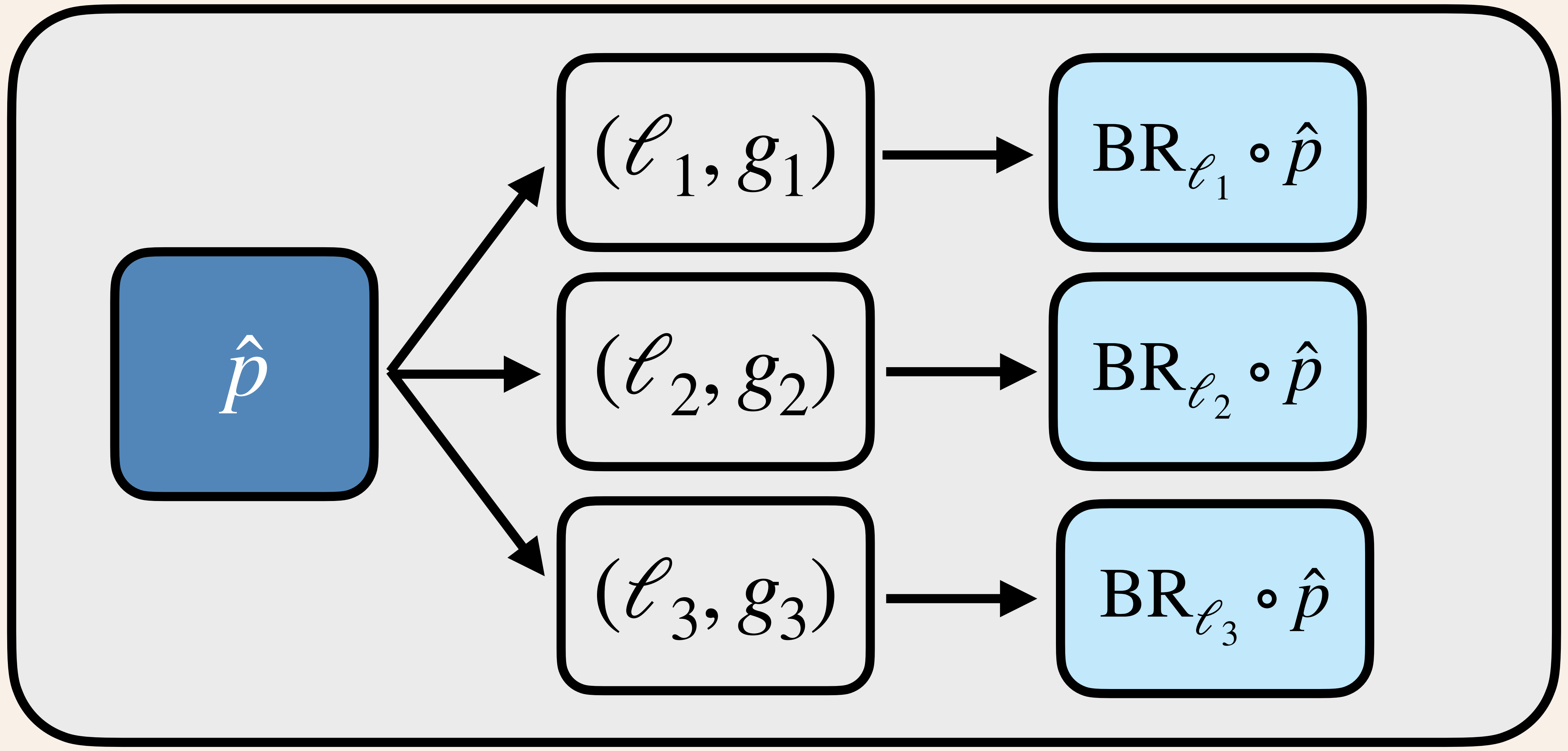
$$\mathop{\mathbb{E}}_{(X,Y)\sim D_g} \left[\ell((\mathrm{BR}_\ell \circ \hat{p})(X), Y)\right] \approx_\varepsilon \min_{h \in H} \mathop{\mathbb{E}}_{(X,Y)\sim D_g} \left[\ell(h(X), Y)\right].$$

Panprediction

Fix $(\mathscr{L}, \mathscr{G}, \mathscr{H})$.

Learn a predictor $\hat{p} : \mathscr{X} \to [0,1]$ such that for all $g \in \mathscr{G}$ and $\ell \in \mathscr{L}$,

$$\mathop{\mathbb{E}}_{(X,Y)\sim D_g}\left[\ell((\mathrm{BR}_\ell \circ \hat{p})(X), Y)\right] \approx_\varepsilon \min_{h\in H} \mathop{\mathbb{E}}_{(X,Y)\sim D_g}\left[\ell(h(X), Y)\right].$$

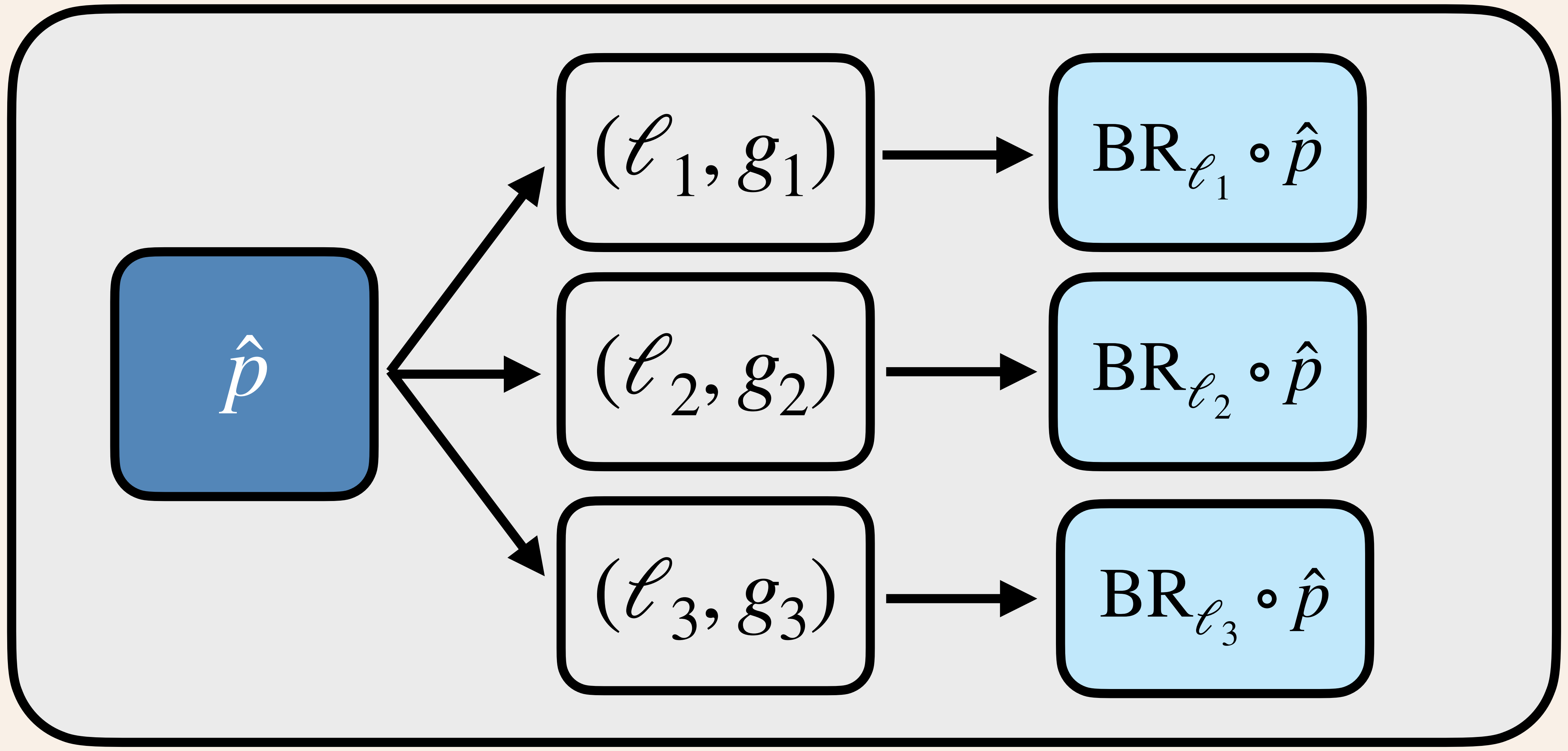


Panprediction

Fix $(\mathscr{L}, \mathscr{G}, \mathscr{H})$.

Learn a predictor $\hat{p} : \mathscr{X} \to [0,1]$ such that for all $g \in \mathscr{G}$ and $\ell \in \mathscr{L}$,

$$\mathop{\mathbb{E}}_{(X,Y)\sim D_g}\left[\ell((\mathrm{BR}_\ell \circ \hat{p})(X), Y)\right] \approx_\varepsilon \min_{h\in H} \mathop{\mathbb{E}}_{(X,Y)\sim D_g}\left[\ell(h(X), Y)\right].$$



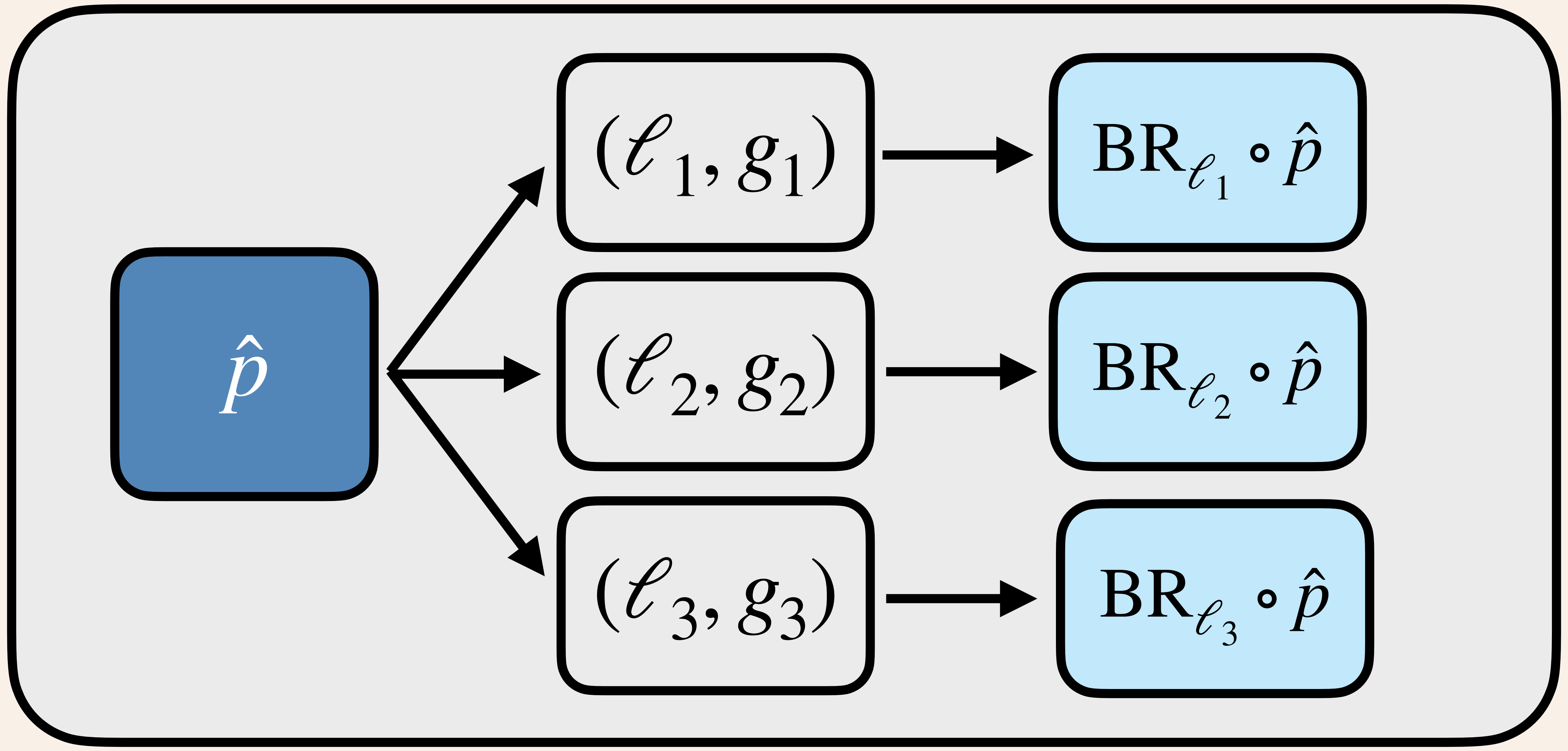
Non-vacuous! **Bayes predictor**
$p^\star(x) = \mathbb{P}_{(X,Y)\sim D}(Y = 1 \mid X = x)$
is panpredictor *for all* $\mathscr{L}, \mathscr{G}, \mathscr{H}$.

Panprediction

Fix $(\mathscr{L}, \mathscr{G}, \mathscr{H})$.

Learn a predictor $\hat{p} : \mathscr{X} \to [0,1]$ such that for all $g \in \mathscr{G}$ and $\ell \in \mathscr{L}$,

$$\mathop{\mathbb{E}}_{(X,Y)\sim D_g} \left[\ell((\mathrm{BR}_\ell \circ \hat{p})(X), Y)\right] \approx_\varepsilon \min_{h\in H} \mathop{\mathbb{E}}_{(X,Y)\sim D_g} \left[\ell(h(X), Y)\right].$$



Non-vacuous! **Bayes predictor** $p^\star(x) = \mathbb{P}_{(X,Y)\sim D}(Y = 1 \mid X = x)$ is panpredictor *for all* $\mathscr{L}, \mathscr{G}, \mathscr{H}$.

Suffices to **learn a coarsening** of $p^\star$

Panprediction $\leq$ Step Calibration

Panprediction $\leq$ Step Calibration

Fix $(\mathscr{L}, \mathscr{G}, \mathscr{H})$.

Learn a predictor $\hat{p} : \mathscr{X} \to [0,1]$ such that for all $v, w \in [0,1]$, $g \in \mathscr{G}$, and $h \in \mathscr{H}$,

$$\left| \mathbb{E}\left[(Y - \hat{p}(X)) \cdot 1\left(\hat{p}(X) \leq v\right) \cdot 1\left(h(X) \leq w\right) \cdot 1\left(g(X) = 1\right) \right] \right| \approx_{\varepsilon} 0$$

Panprediction $\leq$ Step Calibration

Fix $(\mathscr{L}, \mathscr{G}, \mathscr{H})$.

Learn a predictor $\hat{p} : \mathscr{X} \to [0,1]$ such that for all $v, w \in [0,1]$, $g \in \mathscr{G}$, and $h \in \mathscr{H}$,

$$\left| \mathbb{E}\left[(Y - \hat{p}(X)) \cdot 1\left(\hat{p}(X) \leq v\right) \cdot 1\left(h(X) \leq w\right) \cdot 1\left(g(X) = 1\right) \right] \right| \approx_{\varepsilon} 0$$

Panprediction $\leq$ Step Calibration

Fix $(\mathscr{L}, \mathscr{G}, \mathscr{H})$.

Learn a predictor $\hat{p} : \mathscr{X} \to [0,1]$ such that for all $v, w \in [0,1]$, $g \in \mathscr{G}$, and $h \in \mathscr{H}$,

$$\left| \mathbb{E}\left[(Y - \hat{p}(X)) \cdot 1\left(\hat{p}(X) \leq v\right) \cdot 1\left(h(X) \leq w\right) \cdot 1\left(g(X) = 1\right) \right] \right| \approx_{\varepsilon} 0$$

Predictions are **unbiased** over all sets of the form

Sublevel set of $\hat{p}$ $\times$ **Sublevel** set of h $\times$ Level set of g

(Step calibration [QZ25] aka proper calibration [OKK25])

Panprediction $\leq$ Step Calibration

Fix $(\mathscr{L}, \mathscr{G}, \mathscr{H})$.

Learn a predictor $\hat{p} : \mathscr{X} \to [0,1]$ such that for all $v, w \in [0,1]$, $g \in \mathscr{G}$, and $h \in \mathscr{H}$,

$$\left| \mathbb{E}\left[(Y - \hat{p}(X)) \cdot \mathbb{1}\left(\hat{p}(X) \leq v\right) \cdot \mathbb{1}\left(h(X) \leq w\right) \cdot \mathbb{1}\left(g(X) = 1\right) \right] \right| \approx_{\varepsilon} 0$$

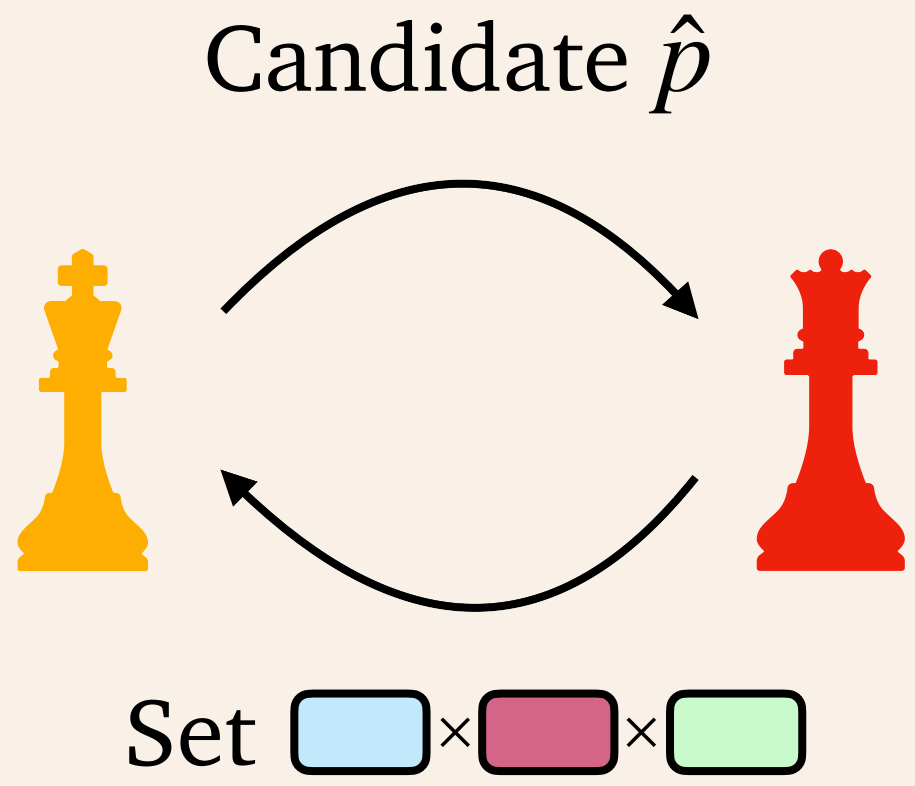
Predictions are **unbiased** over all sets of the form

Sublevel set of $\hat{p}$ $\times$ **Sublevel** set of h $\times$ Level set of g

(Cf. multicalibration, which asks for unbiasedness on products of **level** sets … **inefficient**)

(Step calibration [QZ25] aka proper calibration [OKK25])

Algorithms



Based on **no-regret learning dynamics**

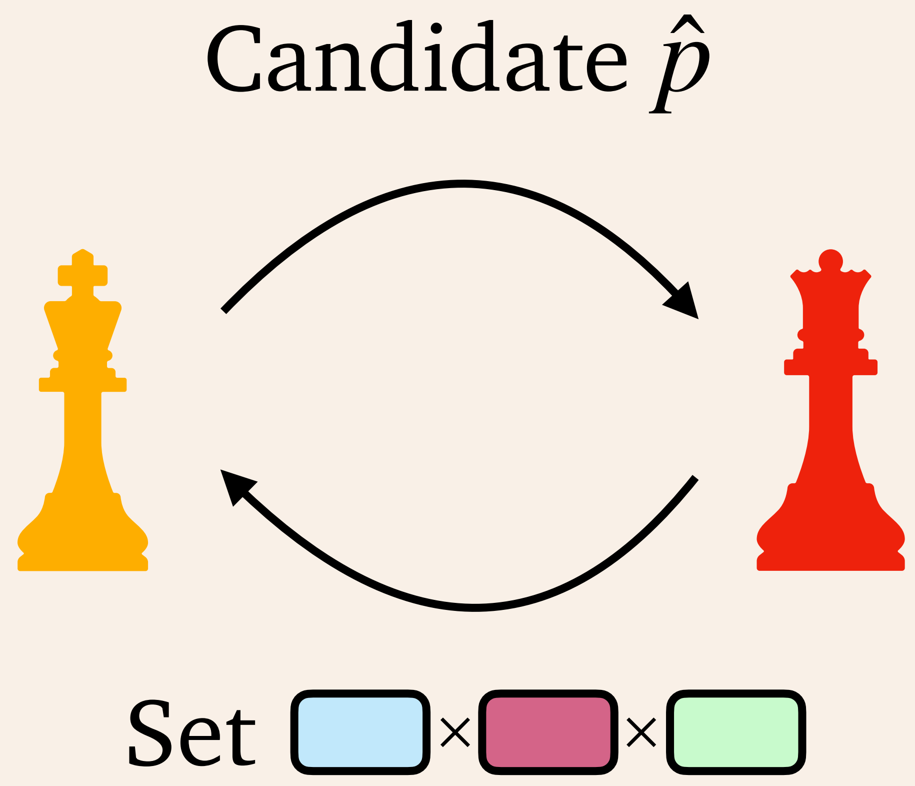
- Primal player: construct candidate panpredictor $\hat{p}$
- Dual player: identify set with large calibration error

Learning panpredictor = finding equilibrium of 2P0S game

Sample complexity $\tilde{O}\left(\frac{\log(|\mathcal{G}||\mathcal{H}|)}{\varepsilon^2}\right)$

(Multi-objective learning and game dynamics [HJZ23])

Algorithms



Based on **no-regret learning dynamics**

- Primal player: construct candidate panpredictor $\hat{p}$
- Dual player: identify set with large calibration error

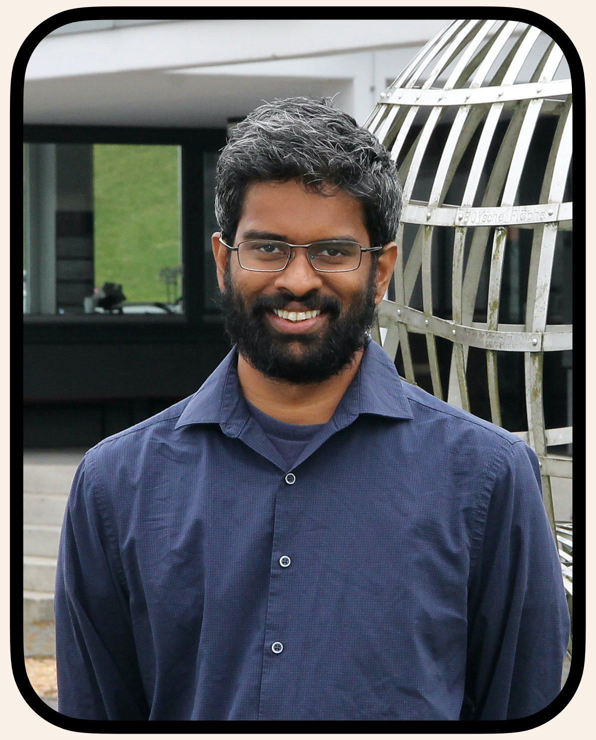
Learning panpredictor = finding equilibrium of 2P0S game

Sample complexity $\tilde{O}\left(\frac{\log(|\mathcal{G}||\mathcal{H}|)}{\varepsilon^2}\right)$

Optimizing **many losses** over **many tasks** can be as **statistically easy** as optimizing one loss over one task, thanks to a reduction to **step calibration**.

(Multi-objective learning and game dynamics [HJZ23])

Thanks!


Siva Balakrishnan


Nika Haghtalab


Daniel Hsu

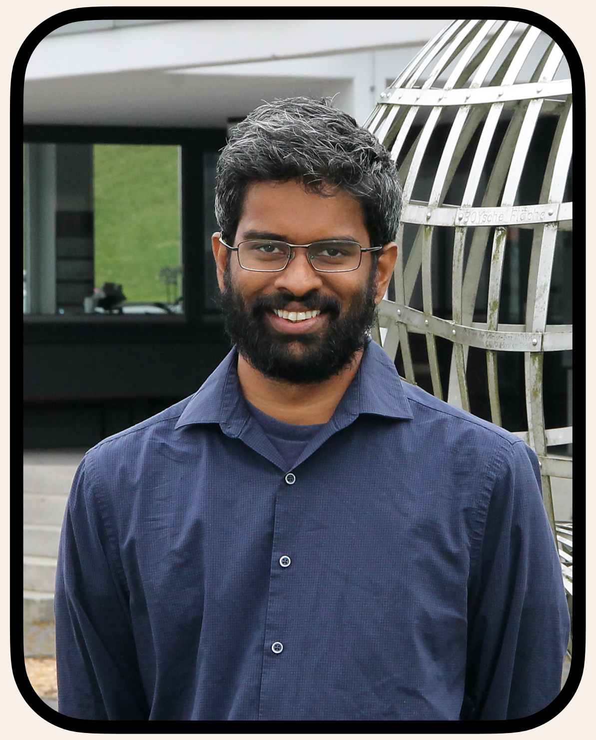

Eric Zhao

References

[RY21] Rothblum and Yona. Multi-group Agnostic PAC Learnability. 2021.
[TH22] Tosh and Hsu. Simple and near-optimal algorithms for multi-group learning. 2022.
[QZ25] Qiao and Zhao. Truthfulness of Decision-theoretic Calibration Measures. 2025.
[GKR+22] Gopalan et al. Omnipredictors. 2022.
[GHR+22] Gopalan et al. Loss Minimization through the Lens of Outcome Indistinguishability. 2022.
[OKK25] Okoroafor, Kleinberg, Kim. Near-Optimal Algorithms for Omniprediction. 2025.
[HJZ23] Haghtalab, Jordan, Zhao. Unifying View of Multicalibration. 2023.

Calibration Workshop
on 05/05

Thanks!



Siva Balakrishnan

Nika Haghtalab

Daniel Hsu

Eric Zhao

References

[RY21] Rothblum and Yona. Multi-group Agnostic PAC Learnability. 2021.

[TH22] Tosh and Hsu. Simple and near-optimal algorithms for multi-group learning. 2022.

[QZ25] Qiao and Zhao. Truthfulness of Decision-theoretic Calibration Measures. 2025.

[GKR+22] Gopalan et al. Omnipredictors. 2022.

[GHR+22] Gopalan et al. Loss Minimization through the Lens of Outcome Indistinguishability. 2022.

[OKK25] Okoroafor, Kleinberg, Kim. Near-Optimal Algorithms for Omniprediction. 2025.

[HJZ23] Haghtalab, Jordan, Zhao. Unifying View of Multicalibration. 2023.