Correct Conditional Forecasts and Proper Scoring Rules in Performative Prediction

Philip Boeken
VU Amsterdam

joint work with Onno Zoeter (Booking.com), Joris M. Mooij (University of Amsterdam)

April 23, 2026

Motivation: predictions that change what they predict

- A Covid-19 incidence forecast triggers lockdowns $\Rightarrow$ changes incidences.
- IPCC temperature forecasts influence policy $\Rightarrow$ influence temperature.
- Traffic forecasts reroute drivers $\Rightarrow$ preempts the traffic jam.

Motivation: predictions that change what they predict

- A Covid-19 incidence forecast triggers lockdowns $\Rightarrow$ changes incidences.
- IPCC temperature forecasts influence policy $\Rightarrow$ influence temperature.
- Traffic forecasts reroute drivers $\Rightarrow$ preempts the traffic jam.

Performative prediction (Perdomo et al., 2020): the forecast is part of the (causal) system.

Clinical ML angle (van Amsterdam et al., 2025): even a *correct* forecast can be harmful if it induces a bad action.

Motivation: predictions that change what they predict

- A Covid-19 incidence forecast triggers lockdowns $\Rightarrow$ changes incidences.
- IPCC temperature forecasts influence policy $\Rightarrow$ influence temperature.
- Traffic forecasts reroute drivers $\Rightarrow$ preempts the traffic jam.

Performative prediction (Perdomo et al., 2020): the forecast is part of the (causal) system.

Clinical ML angle (van Amsterdam et al., 2025): even a *correct* forecast can be harmful if it induces a bad action.

Three questions today:
(i) When do correct (conditional) forecasts exist at all?
(ii) How do we elicit correct forecasts with a scoring rule?
(iii) Does the score align the forecaster with the agent's interests?

Non-existence of correct forecasts: the prophet's dilemma

Forecast $F \in \mathcal{P}_{\mathcal{Y}}$; causal model M with mechanism $P_M(Y \mid \mathrm{do}(F))$.

Definition

F is *correct for* M if $F = P_M(Y \mid \mathrm{do}(F))$.

Non-existence of correct forecasts: the prophet's dilemma

Forecast $F \in \mathcal{P}_{\mathcal{Y}}$; causal model M with mechanism $P_M(Y \mid \mathrm{do}(F))$.

Definition

F is *correct for* M if $F = P_M(Y \mid \mathrm{do}(F))$.

Bad news: correct marginal forecasts need not exist.

Self-defeating prophecy: for $P_0 \neq P_1$,

$$P_M(Y \mid \mathrm{do}(F)) = \begin{cases} P_0 & F = P_1 \\ P_1 & \text{otherwise} \end{cases} \qquad \Rightarrow \qquad F \neq P_M(Y \mid \mathrm{do}(F)) \; \forall F.$$

Non-existence of correct forecasts: the prophet's dilemma

Forecast $F \in \mathcal{P}_{\mathcal{Y}}$; causal model M with mechanism $P_M(Y \mid \mathrm{do}(F))$.

Definition

F is *correct for* M if $F = P_M(Y \mid \mathrm{do}(F))$.

Bad news: correct marginal forecasts need not exist.

Self-defeating prophecy: for $P_0 \neq P_1$,

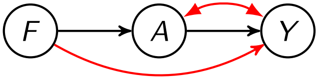
$$P_M(Y \mid \mathrm{do}(F)) = \begin{cases} P_0 & F = P_1 \\ P_1 & \text{otherwise} \end{cases} \qquad \Rightarrow \qquad F \neq P_M(Y \mid \mathrm{do}(F)) \; \forall F.$$

Covid example

Predict few incidences $\Rightarrow$ no lockdown $\Rightarrow$ many incidences.
Predict many incidences $\Rightarrow$ lockdown $\Rightarrow$ few incidences.
No fixed point of the map $F \mapsto P_M(Y \mid \mathrm{do}(F))$.

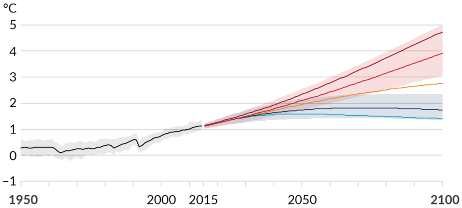
Solution: forecast conditional on a mediator

Conditional forecast. $F(Y \mid A) : \mathcal{A} \to \mathcal{P}_{\mathcal{Y}}$. If A mediates the effect of F on Y:

(∗) 

If $Y \perp\!\!\!\perp_G^d F \mid A$, the target is *forecast-invariant*:

$$P_M(Y \mid A, \mathrm{do}(F)) = P_M(Y \mid A).$$



IPCC: temperature forecasts *given* emission scenario.

Correctness of a conditional forecast

The support of A depends on the forecast F itself, so we distinguish:

Definition

Given $M \in \mathcal{M}_{\mathrm{pc}}$, the conditional forecast $F(Y \mid A)$ is

- *observationally correct* if $F(Y \mid A = a) = P_M(Y \mid A = a, \mathrm{do}(F))$ for every a with $P_M(A = a \mid \mathrm{do}(F)) > 0$;
- *counterfactually correct* if the same holds for every a with $P_M(A = a \mid \mathrm{do}(F)) = 0$;
- *correct* if both.

Characterisation of well-posedness

Theorem (forecastable iff separating)

Let G be a causal graph with variables $F, \mathbf{A}, Y$ such that when all variables in $\mathbf{A}$ are merged into a single variable, this results in a subgraph of $(*)$. Let $\mathcal{M}_G$ be the class of causal models compatible with G. TFAE:

Characterisation of well-posedness

Theorem (forecastable iff separating)

Let G be a causal graph with variables $F, \mathbf{A}, Y$ such that when all variables in $\mathbf{A}$ are merged into a single variable, this results in a subgraph of $(*)$. Let $\mathcal{M}_G$ be the class of causal models compatible with G. TFAE:

1. *For every $M \in \mathcal{M}_G$ there exists an observationally correct forecast of $Y|\mathbf{A}$.*

Characterisation of well-posedness

Theorem (forecastable iff separating)

Let G be a causal graph with variables $F, \mathbf{A}, Y$ such that when all variables in $\mathbf{A}$ are merged into a single variable, this results in a subgraph of $(*)$. Let $\mathcal{M}_G$ be the class of causal models compatible with G. TFAE:

1. *For every $M \in \mathcal{M}_G$ there exists an observationally correct forecast of $Y|\mathbf{A}$.*
2. *For every $M \in \mathcal{M}_G$ there exists a correct forecast of $Y|\mathbf{A}$.*

Characterisation of well-posedness

Theorem (forecastable iff separating)

Let G be a causal graph with variables $F, \mathbf{A}, Y$ such that when all variables in $\mathbf{A}$ are merged into a single variable, this results in a subgraph of $(*)$. Let $\mathcal{M}_G$ be the class of causal models compatible with G. TFAE:

1. *For every $M \in \mathcal{M}_G$ there exists an observationally correct forecast of $Y|\mathbf{A}$.*
2. *For every $M \in \mathcal{M}_G$ there exists a correct forecast of $Y|\mathbf{A}$.*
3. *$Y|\mathbf{A}$ is forecast-invariant for all $M \in \mathcal{M}_G$.*

Characterisation of well-posedness

Theorem (forecastable iff separating)

Let G be a causal graph with variables $F, \mathbf{A}, Y$ such that when all variables in $\mathbf{A}$ are merged into a single variable, this results in a subgraph of $(*)$. Let $\mathcal{M}_G$ be the class of causal models compatible with G. TFAE:

1. *For every $M \in \mathcal{M}_G$ there exists an observationally correct forecast of $Y|\mathbf{A}$.*
2. *For every $M \in \mathcal{M}_G$ there exists a correct forecast of $Y|\mathbf{A}$.*
3. *$Y|\mathbf{A}$ is forecast-invariant for all $M \in \mathcal{M}_G$.*
4. *$Y \perp\!\!\!\perp^d_G F \mid \mathbf{A}$.*

Characterisation of well-posedness

Theorem (forecastable iff separating)

Let G be a causal graph with variables $F, \mathbf{A}, Y$ such that when all variables in $\mathbf{A}$ are merged into a single variable, this results in a subgraph of $(*)$. Let $\mathcal{M}_G$ be the class of causal models compatible with G. TFAE:

1. *For every $M \in \mathcal{M}_G$ there exists an observationally correct forecast of $Y|\mathbf{A}$.*
2. *For every $M \in \mathcal{M}_G$ there exists a correct forecast of $Y|\mathbf{A}$.*
3. *$Y|\mathbf{A}$ is forecast-invariant for all $M \in \mathcal{M}_G$.*
4. *$Y \perp\!\!\!\perp^d_G F \mid \mathbf{A}$.*

Take-away: under graphical assumptions, *conditioning on a separator is both necessary and sufficient* for the forecasting problem to be well-posed.

So correct forecasts exist. Can we *elicit* them?

Proper scoring rules (classical)

Principal pays forecaster $S(F, y)$ after observing y.

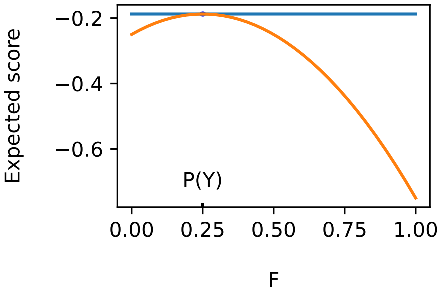
Definition

S is *proper* wrt $\mathcal{P}_\mathcal{Y}$ if

$$\overline{S}(F, P^*) \leq \overline{S}(P^*, P^*) \qquad \forall F, P^* \in \mathcal{P}_\mathcal{Y},$$

strictly proper if equality implies $F = P^*$.

Standard examples: Brier, log-score, energy score, Bregman scores.



$\mathcal{Y} = \{0, 1\}$, $P(Y = 1) = 0.25$,
Brier score: $S(F, y) = -(F - y)^2$.
$\overline{S}(F, P) = -(P - F)^2 - P(1 - P)$ maximised at $F = P$.

Proper scoring rules (classical)

Principal pays forecaster $S(F, y)$ after observing y.

Definition

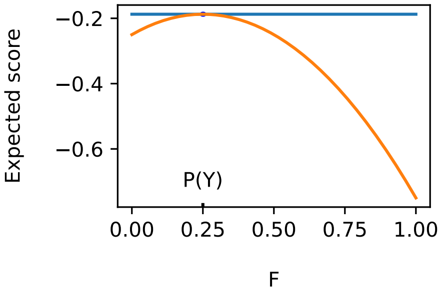
S is *proper* wrt $\mathcal{P}_{\mathcal{Y}}$ if

$$\overline{S}(F, P^*) \leq \overline{S}(P^*, P^*) \qquad \forall F, P^* \in \mathcal{P}_{\mathcal{Y}},$$

strictly proper if equality implies $F = P^*$.

Standard examples: Brier, log-score, energy score, Bregman scores.

Extension to conditional:
$S(F, a, y) := S'(F(Y \mid A = a), y)$.



$\mathcal{Y} = \{0, 1\}$, $P(Y = 1) = 0.25$,
Brier score: $S(F, y) = -(F - y)^2$.
$\overline{S}(F, P) = -(P - F)^2 - P(1 - P)$ maximised at $F = P$.

Proper scoring rules (classical)

Principal pays forecaster $S(F, y)$ after observing y.

Definition

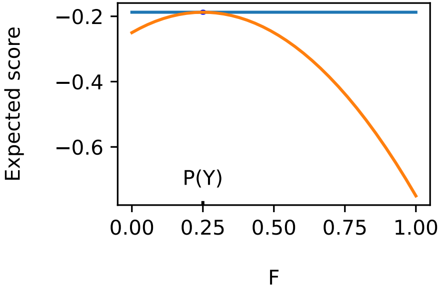
S is *proper* wrt $\mathcal{P}_{\mathcal{Y}}$ if

$$\overline{S}(F, P^*) \leq \overline{S}(P^*, P^*) \qquad \forall F, P^* \in \mathcal{P}_{\mathcal{Y}},$$

strictly proper if equality implies $F = P^*$.

Standard examples: Brier, log-score, energy score, Bregman scores.

Extension to conditional:
$S(F, a, y) := S'(F(Y \mid A = a), y)$.



$\mathcal{Y} = \{0, 1\}$, $P(Y = 1) = 0.25$,
Brier score: $S(F, y) = -(F - y)^2$.
$\overline{S}(F, P) = -(P - F)^2 - P(1 - P)$ maximised at $F = P$.

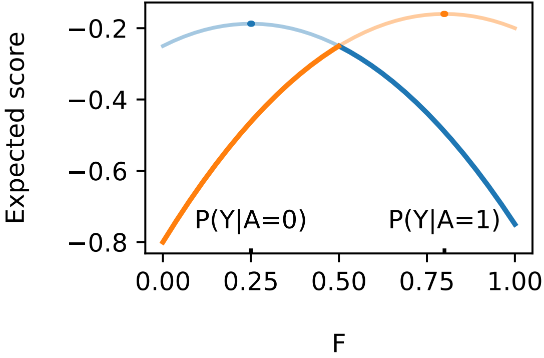
Question. Under forecast-invariance, does an S that's strictly proper *classically* elicit correct *performative* forecasts?

Self-defeating prophecy: nothing to elicit

$\mathcal{Y} = \{0, 1\}$, marginal forecast $F \in [0, 1]$, Brier score.

$$P(Y = 1 \mid A) = \begin{cases} 0.8 & A = 0 \\ 0.25 & A = 1. \end{cases}$$

$$A = \begin{cases} 1 & F > 0.5 \text{ (intervene)} \\ 0 & \text{otherwise} \end{cases}$$

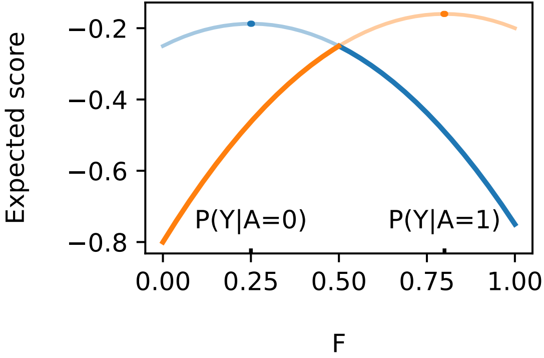


Self-defeating prophecy: nothing to elicit

$\mathcal{Y} = \{0, 1\}$, marginal forecast $F \in [0, 1]$, Brier score.

$$P(Y = 1 \mid A) = \begin{cases} 0.8 & A = 0 \\ 0.25 & A = 1. \end{cases}$$

$$A = \begin{cases} 1 & F > 0.5 \text{ (intervene)} \\ 0 & \text{otherwise} \end{cases}$$



No correct forecast exists:

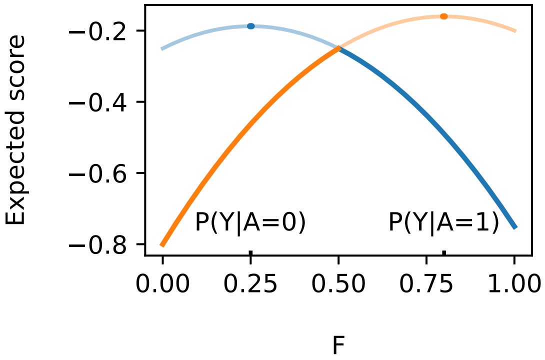
- $F > 0.5$: truth is 0.25, need $F = 0.25$. $\times$
- $F \leq 0.5$: truth is 0.8, need $F = 0.8$. $\times$

Self-defeating prophecy: nothing to elicit

$\mathcal{Y} = \{0, 1\}$, marginal forecast $F \in [0, 1]$, Brier score.

$$P(Y = 1 \mid A) = \begin{cases} 0.8 & A = 0 \\ 0.25 & A = 1. \end{cases}$$

$$A = \begin{cases} 1 & F > 0.5 \text{ (intervene)} \\ 0 & \text{otherwise} \end{cases}$$



No correct forecast exists:

- $F > 0.5$: truth is 0.25, need $F = 0.25$. ×
- $F \leq 0.5$: truth is 0.8, need $F = 0.8$. ×

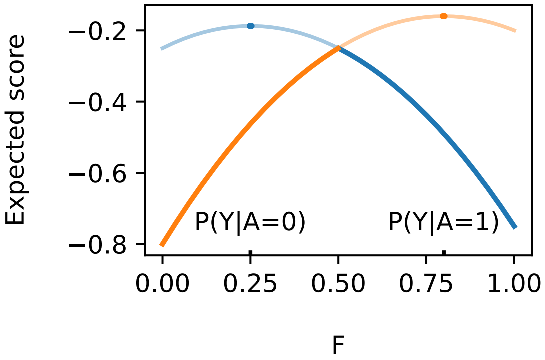
With no correct forecast, **properness is ill-defined**.

Self-defeating prophecy: nothing to elicit

$\mathcal{Y} = \{0, 1\}$, marginal forecast $F \in [0, 1]$, Brier score.

$$P(Y = 1 \mid A) = \begin{cases} 0.8 & A = 0 \\ 0.25 & A = 1. \end{cases}$$

$$A = \begin{cases} 1 & F > 0.5 \text{ (intervene)} \\ 0 & \text{otherwise} \end{cases}$$



No correct forecast exists:

- $F > 0.5$: truth is 0.25, need $F = 0.25$. ×
- $F \leq 0.5$: truth is 0.8, need $F = 0.8$. ×

With no correct forecast, **properness is ill-defined**.

Idea: conditioning on the variable A restores existence (Thm 1).
Properness then needs a richer definition, since the support of A depends on F.

Defining properness for performative conditional forecasts

Expected score for prediction F in causal model M:

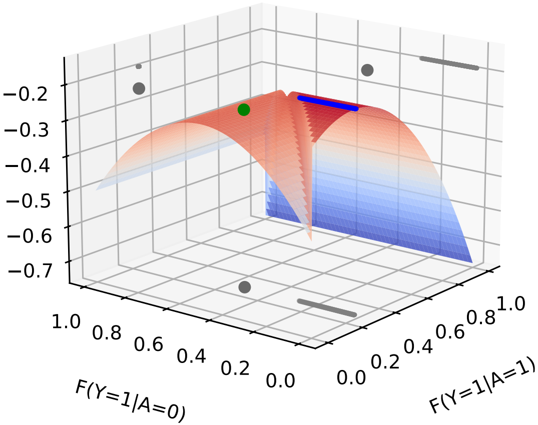
$$\overline{S}_{\mathrm{pc}}(F, M) := \int S(F, a, y)\, P_M(\mathrm{d}a, \mathrm{d}y \mid \mathrm{do}(F)).$$

Wit respect to class of models $\mathcal{M}$, a conditional scoring rule is:

- *observationally (strictly) proper* if $\overline{S}_{\mathrm{pc}}$ is (only) maximised at observationally correct forecasts for all $M \in \mathcal{M}$;
- *counterfactually (strictly) proper* if $\overline{S}_{\mathrm{pc}}$ is (only) maximised at counterfactually correct forecasts for all $M \in \mathcal{M}$;
- *(strictly) proper* if both hold.

Pathology 1: observationally proper, counterfactually not

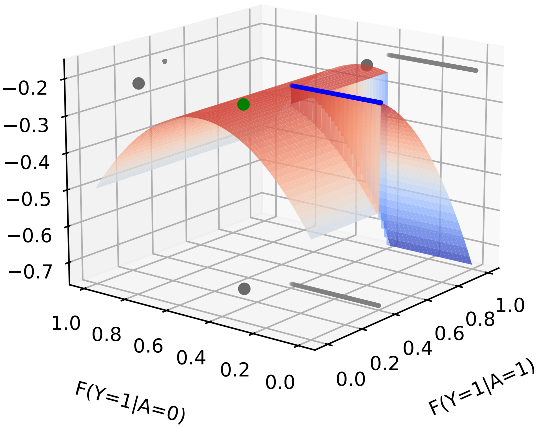
Decision rule $A = \text{argmax}_a F(Y = 1 \mid A = a)$, Brier score.



- Correct forecast (green) is optimal for $F(Y|A = 1)$ (the action that gets observed).
- But optimal surface is flat/decreasing in $F(Y|A = 0)$: the principal can't tell a correct counterfactual from a wrong one.

Intuition. A classical strictly proper score incentivises the forecaster to *manipulate towards the action under which Y is easiest to predict* — here, $A = 1$ since $P(Y = 1|A = 1) = 0.25$ (easy) and $P(Y = 1|A = 0) = 0.5$ (hard)

Pathology 2: neither observationally nor counterfactually proper



- Correct forecast $(0.5, 0.25)$ induces $A = 0$ (hard): score -0.25.
- Near-correct $(0.4, 0.4)$ trips the rule $\Rightarrow A = 1$ (easier): score -0.21.
- Optimum sits *off* the correct forecast in *both* branches.

Intuition. Self-defeating prophecy at work: whenever the forecaster gets close to the truth in the harder regime, the mechanism flips to the easier one.

Impossibility theorem

Theorem (non-existence of proper classical scoring)

Let $\mathcal{M} \subseteq \mathcal{M}_{\mathrm{pc}}$ be models with $Y|A$ forecast-invariant and either full-support or deterministic $P_M(A \mid \mathrm{do}(F))$.

If S admits $P_0, P_1, F \in \mathcal{P}_{\mathcal{Y}}$ with

$$\overline{S}(P_0, P_0) < \overline{S}(F, P_1) < \overline{S}(P_1, P_1),$$

then S is **not** *observationally proper w.r.t. $\mathcal{M}$.*

Impossibility theorem

Theorem (non-existence of proper classical scoring)

Let $\mathcal{M} \subseteq \mathcal{M}_{\mathrm{pc}}$ be models with $Y|A$ forecast-invariant and either full-support or deterministic $P_M(A \mid \mathrm{do}(F))$.

If S admits $P_0, P_1, F \in \mathcal{P}_{\mathcal{Y}}$ with

$$\overline{S}(P_0, P_0) < \overline{S}(F, P_1) < \overline{S}(P_1, P_1),$$

then S is **not** *observationally proper w.r.t. $\mathcal{M}$.*

Translation: any classically proper S with a non-trivial entropy gap (i.e. some distributions are easier to forecast than others) is not proper under performativity. Even if we make conditional forecasts.

Impossibility theorem

Theorem (non-existence of proper classical scoring)

Let $\mathcal{M} \subseteq \mathcal{M}_{\mathrm{pc}}$ be models with $Y|A$ forecast-invariant and either full-support or deterministic $P_M(A \mid \mathrm{do}(F))$.

If S admits $P_0, P_1, F \in \mathcal{P}_\mathcal{Y}$ with

$$\overline{S}(P_0, P_0) < \overline{S}(F, P_1) < \overline{S}(P_1, P_1),$$

then S is **not** *observationally proper w.r.t. $\mathcal{M}$.*

Translation: any classically proper S with a non-trivial entropy gap (i.e. some distributions are easier to forecast than others) is not proper under performativity. Even if we make conditional forecasts.

Two solutions / scenarios:

- decision theoretic setting
- divergence scoring

Solution 1: decision theory — utility scores

Agent with utility $U(a, y)$ picks $a_F := \text{argmax}_a \int U(a, y)\, F(\mathrm{d}y \mid a)$, so $P_M(A \mid \text{do}(F)) = \delta_{a_F}$.

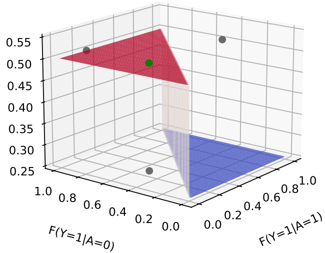
Solution 1: decision theory — utility scores

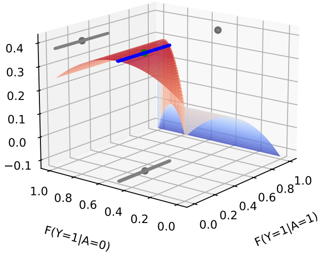
Agent with utility $U(a, y)$ picks $a_F := \text{argmax}_a \int U(a, y)\, F(\mathrm{d}y \mid a)$, so $P_M(A \mid \text{do}(F)) = \delta_{a_F}$.

Theorem (utility scores)

Utility score *(Oesterheld and Conitzer, 2020): $S(F, y) := U(a_F, y)$ is proper, but not observationally strictly proper.*
Strictly proper version: *for $U, S' \in [0, 1]$, S' classically strictly proper, 'utility gap' $> \Delta$,*

$$S(F, y) := U(a_F, y) + \Delta \cdot S'(F(Y \mid A = a_F), y) \qquad \textit{is obs. strictly proper, cf. proper .}$$





Are these scores aligned with the agent's interests?

Definition

Score S is *incentive-compatible* with utility U on $\mathcal{M}$ if

$$\operatorname*{argmax}_F \overline{S}_{\mathrm{pc}}(F, M) \;=\; \operatorname*{argmax}_F \int U(a_F, y)\, P_M(\mathrm{d}y \mid \mathrm{do}(F)) \qquad \forall M \in \mathcal{M}.$$

Are these scores aligned with the agent's interests?

Definition

Score S is *incentive-compatible* with utility U on $\mathcal{M}$ if

$$\operatorname*{argmax}_F \overline{S}_{\mathrm{pc}}(F, M) \;=\; \operatorname*{argmax}_F \int U(a_F, y)\, P_M(\mathrm{d}y \mid \mathrm{do}(F)) \qquad \forall M \in \mathcal{M}.$$

1. Strict properness ⇒ incentive compatibility: the unique argmax is the truth, which induces the optimal action a^*.
2. But strict properness is **impossible**: the score only sees $F(Y \mid A = a_F)$, so counterfactual branches $F(Y \mid A = a')$ cannot be uniquely pinned down.
3. Resort to *observational* strict properness — which does **not** imply incentive compatibility in general: an obs. correct forecast can still induce a harmful action (cf. Pathology 1) (van Amsterdam et al., 2025).

Are these scores aligned with the agent's interests?

Definition

Score S is *incentive-compatible* with utility U on $\mathcal{M}$ if

$$\operatorname*{argmax}_F \overline{S}_{\mathrm{pc}}(F, M) \;=\; \operatorname*{argmax}_F \int U(a_F, y)\, P_M(\mathrm{d}y \mid \mathrm{do}(F)) \qquad \forall M \in \mathcal{M}.$$

1. Strict properness $\Rightarrow$ incentive compatibility: the unique argmax is the truth, which induces the optimal action a^*.
2. But strict properness is **impossible**: the score only sees $F(Y \mid A = a_F)$, so counterfactual branches $F(Y \mid A = a')$ cannot be uniquely pinned down.
3. Resort to *observational* strict properness — which does **not** imply incentive compatibility in general: an obs. correct forecast can still induce a harmful action (cf. Pathology 1) (van Amsterdam et al., 2025).

Theorem

The utility scores above are incentive-compatible and observationally (strictly) proper and counterfactually proper, by construction.

Solution II: score with the performative divergence

$$D(F, P) := \int (S(P, y) - S(F, y))\, P(\mathrm{d}y) \quad \geq 0, \; = 0 \iff F = P \text{ (if } S \text{ str. proper).}$$

Definition

Performative divergence:

$$D_{\mathrm{pc}}(F, M) := \int (S(P_M(Y \mid A = a, \mathrm{do}(F)), y) - S(F(Y \mid A = a), y))\, P_M(\mathrm{d}a, \mathrm{d}y \mid \mathrm{do}(F)).$$

Solution II: score with the performative divergence

$$D(F,P) := \int (S(P,y) - S(F,y))\, P(\mathrm{d}y) \quad \geq 0, \; = 0 \iff F = P \text{ (if } S \text{ str. proper).}$$

Definition

Performative divergence:

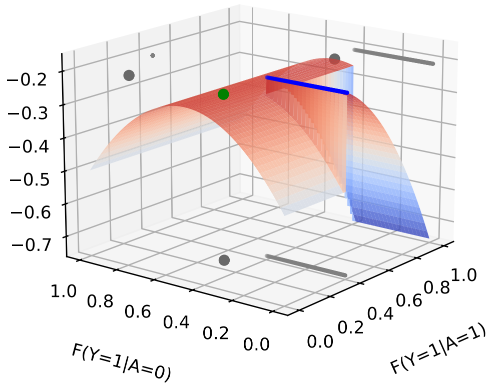
$$D_{\mathrm{pc}}(F,M) := \int (S(P_M(Y \mid A = a, \mathrm{do}(F)), y) - S(F(Y \mid A = a), y))\, P_M(\mathrm{d}a, \mathrm{d}y \mid \mathrm{do}(F)).$$

Theorem (divergence is proper)

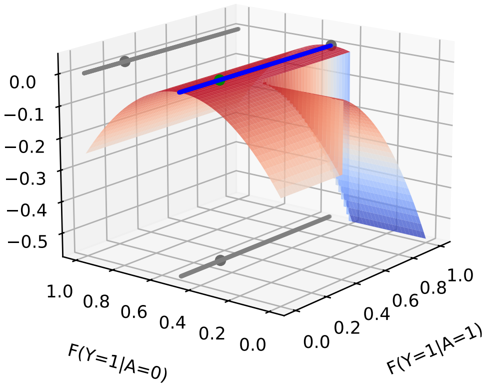
If S is (strictly) proper classically, then D_{pc} is observationally (strictly) proper and counterfactually proper on any $\mathcal{M}'$ where for each model a correct forecast exist.

Divergence on the running example

Same setup as Pathology 2 (self-defeating prophecy).



Brier score — improper



Brier divergence — obs. strictly proper

The correct forecast (green) is the unique maximiser of the divergence — even without positivity, and without any change to the mechanism.

Estimating the divergence

Downside: divergence requires the principal to *estimate* D_{pc}.

Corollary

Any unbiased estimator $\hat{D}_{\mathrm{pc}}$ *of* $D_{\mathrm{pc}}(F, M)$ *is observationally (strictly) proper — for any sample size* n *for which it is well-defined, and regardless of its variance.*

Estimating the divergence

Downside: divergence requires the principal to *estimate* D_{pc}.

Corollary

Any unbiased estimator $\hat{D}_{\mathrm{pc}}$ *of* $D_{\mathrm{pc}}(F, M)$ *is observationally (strictly) proper — for any sample size* n *for which it is well-defined, and regardless of its variance.*

Examples. Given i.i.d. observations $(A_1, Y_1), \ldots, (A_n, Y_n) \sim P_M(A, Y \mid \mathrm{do}(F))$, write $n_a := \sum_i \mathbb{1}\{A_i = a\}$, $\hat{p}_a := \frac{1}{n_a} \sum_i Y_i \mathbb{1}\{A_i = a\}$, $f_a := F(Y = 1 \mid A = a)$.

- Brier score: $\hat{D}^{\mathrm{Brier}} = \frac{1}{n} \sum_i (\hat{p}_{a_i} - f_{a_i})^2 - \dfrac{\hat{p}_{a_i}(1 - \hat{p}_{a_i})}{n_{a_i} - 1}$.
- Energy score: similar bias-corrected form (see paper).

Needs $n_a \geq 2$ for each a in the support of $P_M(A \mid \mathrm{do}(F))$.

Estimating forecast parameters

Instead of taking the perspective of the principal, switch to the perspective of the forecaster. How to estimate 'correct' parameters if the data is affected by previous model?

Perdomo et al. (2020): in performative settings, ordinary ERM can diverge.

- *Performative risk:* $R^{\mathrm{p}}(\theta) := \int \ell(\theta, a, y) \, P_M(\mathrm{d}a, \mathrm{d}y \,|\, \mathrm{do}(F_\theta))$.
- *Decoupled risk:* $R^{\mathrm{d}}(\theta_{t+1}, \theta_t) := \int \ell(\theta_{t+1}, a, y) \, P_M(\mathrm{d}a, \mathrm{d}y \,|\, \mathrm{do}(F_{\theta_t}))$.
- *Performatively stable:* $\theta_{PS} = \mathrm{argmin}_\theta \, R^{\mathrm{d}}(\theta, \theta_{PS})$.
- *Performatively optimal:* $\theta_{PO} \in \mathrm{argmin}_\theta \, R^{\mathrm{p}}(\theta)$.

Estimating forecast parameters

Instead of taking the perspective of the principal, switch to the perspective of the forecaster. How to estimate 'correct' parameters if the data is affected by previous model?

Perdomo et al. (2020): in performative settings, ordinary ERM can diverge.

- *Performative risk:* $R^{\mathrm{p}}(\theta) := \int \ell(\theta, a, y)\, P_M(\mathrm{d}a, \mathrm{d}y \,|\, \mathrm{do}(F_\theta))$.
- *Decoupled risk:* $R^{\mathrm{d}}(\theta_{t+1}, \theta_t) := \int \ell(\theta_{t+1}, a, y)\, P_M(\mathrm{d}a, \mathrm{d}y \,|\, \mathrm{do}(F_{\theta_t}))$.
- *Performatively stable:* $\theta_{PS} = \mathrm{argmin}_\theta\, R^{\mathrm{d}}(\theta, \theta_{PS})$.
- *Performatively optimal:* $\theta_{PO} \in \mathrm{argmin}_\theta\, R^{\mathrm{p}}(\theta)$.

Theorem

Let $Y|A$ be forecast-invariant, $P_M(A \,|\, \mathrm{do}(F))$ have full support, and D the divergence of a strictly proper S. For any θ_t, the parameter

$$\theta_{t+1} := \underset{\theta}{\mathrm{argmin}}\, R^{\mathrm{d}}_D(\theta, \theta_t)$$

yields a correct forecast which is **performatively stable** *and* **performatively optimal***.*

Take-aways

- Performative forecasts can lack a fixed point.
- Making conditional forecasts restores well-posedness.
- Forecast-invariance ($Y \perp\!\!\!\perp^{d} F \mid A$) is the **characterising** condition.
- Classical strictly proper scoring rules are **not** performatively proper — even under forecast-invariance.
- Two fixes:
  - **Decision theory:**
    - utility score (proper)
    - utility score + Δ-scaled proper score (obs. strictly proper, cf. proper)

    Auxiliary result: these scores are incentive-compatible.
  - **Divergence:**
    - divergence D_{pc} is obs. str. proper
    - unbiased estimate $\hat{D}_{\mathrm{pc}}$ is obs. str. proper as well.
- Training with the divergence $\Rightarrow$ single-step performative stability *and* optimality.

Thank you!

References I

Oesterheld, C. and Conitzer, V. (2020). Decision Scoring Rules. In *Web and Internet Economics: 16th International Conference, WINE 2020*.

Perdomo, J., Zrnic, T., Mendler-Dünner, C., and Hardt, M. (2020). Performative Prediction. In *Proceedings of the 37th International Conference on Machine Learning*, pages 7599–7609. PMLR.

van Amsterdam, W. A. C., van Geloven, N., Krijthe, J. H., Ranganath, R., and Cinà, G. (2025). When accurate prediction models yield harmful self-fulfilling prophecies. *Patterns (New York, N.Y.)*, 6(4):101229.