Prior-Free Information Design

Maxwell Rosenthal*

November 25, 2025

Abstract

This paper introduces a prior-free framework for information design based on partial identification and applies it to robust causal inference. The decision maker observes the distribution of signals generated by an information structure and ranks alternatives by their worst-case payoff over the state distributions consistent with those signals. We characterize the set of robustly implementable actions and show that each can be implemented by an information structure that withholds at most one dimension of information from the decision maker. In the potential outcomes model, every treatment is implementable via an experiment that is almost fully informative.

1 Introduction

Frequentist data analysis is standard in regulatory evaluations of medical interventions and in scientific publishing. At the same time, existing models of strategic communication typically study Bayesian decision makers who filter evidence through an exogenous prior. Empirical researchers are themselves economic agents who face private incentives and might therefore be tempted to selectively disclose information, even when what they do disclose must accurately reflect their data. How should an objective decision maker respond?

In our model, the decision maker observes the distribution of signals generated by an information structure. He understands the stochastic relationship between states

*Georgia Institute of Technology. Email: rosenthal@gatech.edu.

and signals, and views any distribution of states that rationalizes the distribution of signals as plausible. The decision maker's payoffs are state-dependent; he ranks actions according to their worst-case payoff over the partially identified set of plausible state distributions; and the information structure *implements* an action if it is worst-case optimal under that structure. Our model is directly applicable to causal inference.

Consider the classical treatment effects environment with binary outcome Y, binary treatment T, covariate X, and potential outcomes (Y_0, Y_1). Treatment is unconfounded, the assignment mechanism $P(T \mid X)$ is known, and the joint distribution of (Y, X, T) identifies $\mathbb{E}[Y_0]$ and $\mathbb{E}[Y_1]$. In contrast, if the decision maker observes only the marginal distribution of (Y, T), then uncertainty about how potential outcomes vary with X translates into uncertainty about the consequences of each policy. There are situations in which a fully informed decision maker declines the treatment while a partially informed decision maker accepts it.

This paper's general results are as follows. Action α is implementable if and only if there exists a *supporting prior* ν under which the expected utility of α (i) is greater than or equal to that of any other action β and (ii) does not exceed the expected utility of α with respect to the true state distribution μ. All such actions are implementable via *almost fully informative* information structures under which the decision maker faces a single dimension of uncertainty. We apply this result to the general causal inference framework with many treatments, many covariates, and finitely many outcomes. In that special case of our model, every available treatment is implementable via an information structure that is strictly more informative than disclosure of the marginal distribution of any proper subset of the observed variables.

Our decision maker behaves as if he knew the state variable was distributed according to some adversarial prior ν in the identified set. In order for an information structure (Σ, E) to implement its intended action α, two things must be true. First, α must maximize expected utility with respect to ν. Second, the decision maker's payoff for α under ν must be no better than under the true distribution μ. As we show, these weak necessary conditions are also sufficient. An action is implementable if and only if there exists a distribution of the state variable that satisfies them.

Related literature This paper endogenizes partial identification in the same way Bayesian persuasion endogenizes Blackwell experiments. In doing so, we contribute to the literatures on experimentation and information design, robust decision making in economic theory, and partial identification in econometrics.

Blackwell's seminal papers (1951; 1953) introduce a model of statistical experimentation in which a Bayesian decision maker observes a signal generated by the experiment and updates his beliefs before acting. Kamenica and Gentzkow (2011) endogenize the choice of experiment in a sender–receiver environment, characterize the sender–optimal experiment, and thereby launch the Bayesian persuasion literature. In turn, Bergemann and Morris (2016) introduce the Bayes correlated equilibrium solution concept as a general framework for information design in Bayesian games with multiple agents, and survey the literature in Bergemann and Morris (2019). In contrast, our decision maker responds to the entire distribution of signals rather than realization-by-realization, has no prior, and evaluates actions by their worst-case payoff over the identified set.

Our model of individual decision making and our ranking of information structures were originally developed in an earlier paper Rosenthal (2025). There, information structure (Σ, E) is *robustly more informative* than (Σ', E') if for every decision problem the decision maker's guaranteed payoff under the former matches or exceeds his guaranteed payoff under the latter. This order is implied by, but does not imply, Blackwell's classical order. Outside of this framework, there is a small but growing literature on non-Bayesian decision making in Blackwell experimentation and sender–receiver games. Whitmeyer (2025) studies the information–monotonicity of non-Bayesian updating rules, and Yang, Yoder, and Zentefis (2025) study worst-case decision making under finite-dimensional misspecification of infinite-dimensional models. An earlier stream of papers extends Blackwell's classical framework to maxmin expected-utility maximizers who make Bayesian updates to exogenously specified sets of priors (Çelen (2012); Heyen and Wiesenfarth (2015); Li and Zhou (2016)).

Finally, the partial-identification literature in econometrics initiated in (Manski, 1990, 1997, 2003, 2007, 2013) and surveyed by Tamer (2010); Molinari (2020); Kline and Tamer (2023) studies decision problems in which the distribution of the state is set-identified and actions are evaluated according to worst-case criteria. We are distinguished from this literature by our endogenization of the identified set via our interpretation of the information structure as a choice variable. This treatment facilitates our characterization of the actions that can be implemented by some information structure and our identification of the maximally informative structures that do so.

The paper is organized as follows. In Section 2, we illustrate with a numerical treatment effects example. We describe the formal model in Section 3, state our main results in Section 4, formally develop our application to causal inference in Section 5, and conclude in Section 6. Proofs and other technical material are in the Appendix.

2 Motivating Example

In our causal-inference application, there exist data sets for which partial disclosure incentivizes actions that full disclosure does not. A researcher discloses population-level data about the effect of binary treatment T to a decision maker. The untreated outcome Y_0, the treated outcome Y_1, and the covariate X are binary. The researcher observes the true distribution μ of (Y, X, T) presented in Table 1, treatment is unconfounded, and the assignment mechanism

$$P(T = 1 \mid X = 0) = 0.2, \qquad\qquad P(T = 1 \mid X = 1) = 0.8$$

is common knowledge.

Table 1: Observed distribution $\mu(Y, X, T)$

	$(X{=}0, T{=}0)$	$(X{=}1, T{=}0)$	$(X{=}0, T{=}1)$	$(X{=}1, T{=}1)$
$Y = 0$	0.40	0.05	0.10	0.30
$Y = 1$	0.00	0.05	0.00	0.10

The decision maker's problem is to choose a treatment $a \in \{0, 1\}$ to maximize expected outcomes $\mathbb{E}[Y_a]$. If the researcher discloses full joint distribution μ of the data, then the counterfactual means

$$E[Y_0] = \sum \frac{\mathbf{1}\{T = 0\}Y}{P(T = 0|X)}\mu(Y, T, X) = 0.25,$$
$$E[Y_1] = \sum \frac{\mathbf{1}\{T = 1\}Y}{P(T = 1|X)}\mu(Y, T, X) = 0.125$$

are fully identified and the decision maker declines to implement the treatment.

Suppose instead that the researcher discloses only the marginal distribution of (Y, T). The decision maker evaluates treatment a by the worst-case payoff $\mathbb{E}_\nu[Y_a]$ with respect to all distributions ν on (Y, X, T) consistent with the known assignment mechanism and the disclosed marginal in Table 2.

Table 2: Disclosed marginal distribution $\mu(Y, T)$

	$T = 0$	$T = 1$
$Y = 0$	0.45	0.40
$Y = 1$	0.05	0.10

The worst case for $a = 0$ concentrates as many of the decision maker's preferred $Y = 1$ outcomes as possible in the $X = 0$ group with $P(T = 1 \mid X = 0) = 0.2$, and the worst case for $a = 1$ concentrates as many of those outcomes as possible in the $X = 1$ group with $P(T = 1 \mid X = 1) = 0.8$. Intuitively, if the data (Y, T) were generated by a process that assigns treatment T to X-groups that disproportionately benefit from T , then assigning treatment to the entire population is less effective than it is under other joint distributions with the disclosed marginal. In our decision maker's case, the joint distribution ν^* in Table 3 is worst-case for both actions, and his payoffs are

$$\mathbb{E}_{\nu^*}[Y_0] = 0.0625, \qquad \mathbb{E}_{\nu^*}[Y_1] = 0.125.$$

While treatment is suboptimal when the decision maker observes the joint distribution of (Y, X, T), it is robustly optimal when he observes only (Y, T). Absent bias toward a no-treatment status quo, partial disclosure reverses the full-information optimal policy.

Table 3: Worst-case joint distribution $\nu^*(Y, X, T)$ consistent with $\mu(Y, T)$

	$(X{=}0, T{=}0)$	$(X{=}1, T{=}0)$	$(X{=}0, T{=}1)$	$(X{=}1, T{=}1)$
$Y = 0$	0.35	0.10	0.10	0.30
$Y = 1$	0.05	0.00	0.00	0.10

3 Model

We write $\mathbb{R}^n$ for the set of real vectors of length n equipped with the standard metric, use the notation $\langle x, y\rangle \equiv \sum_i x_i y_i$ for the usual inner product, and identify real valued functions on finite sets X with vectors in $\mathbb{R}^{|X|}$. Given a finite set X, we write $\Delta(X)$ for the set of all probabilities on X and interpret the elements of $\Delta(X)$ as real vectors. More broadly, given a metric space X, we give the set of Borel distributions $\Delta(X)$ the usual weak topology and write $\mathrm{supp}(\nu)$ for the support of $\nu \in \Delta(X)$.

This paper makes extensive use of standard results from linear algebra. Given a set of real vectors $S \subset \mathbb{R}^n$, we write $\mathrm{span}(S)$ for its span and $\mathrm{conv}(S)$ for its convex hull. If S is singleton we write $\mathrm{span}\{v\}$ to improve readability; if S is a vector space we write $\dim(S)$ for its dimension. Given a linear map $f : \mathbb{R}^n \to \mathbb{R}^m$, we write $\ker f \equiv \{v \in \mathbb{R}^n \mid f(v) = 0\}$ and d for a generic element of $\ker f$.

States, actions, and beliefs The set Ω of states is finite and the set A of actions is a non-empty and compact metric space. The set $\mathcal{P} \subset \Delta(\Omega)$ of priors is non-empty, compact, and convex; the state of the world is distributed according to $\mu \in \mathcal{P}$; and the decision maker's utility $u : A \times \Omega \to \mathbb{R}$ depends continuously on both his action and the state. We write ω for a generic element of Ω, a for a generic element of A, and α for a generic element of $\Delta(A)$. Given action $\alpha \in \Delta(A)$ and prior $\nu \in \mathcal{P}$ we write

$$\langle \alpha, \nu \rangle \equiv \int_A \langle a, \nu \rangle \, \mathrm{d}\alpha(a)$$

for the decision maker's expected payoff.

Information structures In our model, the decision maker does not know the distribution of the state of the world. Instead, he makes inferences about μ via *information structure* (Σ, E), where Σ is a finite set of messages and *experiment* $E : \Omega \mapsto \Delta(\Sigma)$ assigns message distributions $E(\cdot \mid \omega)$ to states ω. The decision maker observes the entire distribution

$$(E\mu)(\sigma) \equiv \sum_{\omega \in \Omega} E(\sigma|\omega)\mu(\omega)$$

of messages generated by the information structure and views prior ν in $\mathcal{P}$ as plausible if and only if $E\nu = E\mu$. We write

$$\mathcal{P}(\Sigma, E) \equiv \{\nu \in \mathcal{P} | E\nu = E\mu\} = \{\nu \in \mathcal{P} | (\nu - \mu) \in \ker E\}$$

for the set of all such distributions. Information structure (Σ, E) is *fully informative* if $\ker E = \{0\}$ and *almost fully informative* if $\ker E$ has dimension at most 1, noting that the former is a special case of the latter.

Assumption 1. *For each state ω there exists a distinct state ω' such that (i) $u(a,\omega) = u(a,\omega')$ for all actions $a \in A$ and (ii) if $\nu(s) = \nu'(s)$ for all $s \in \Omega \setminus \{\omega, \omega'\}$ then $\nu \in \mathcal{P}$ if and only if $\nu' \in \mathcal{P}$.*

Assumption 1 enriches the state space with payoff-irrelevant variables. These signals allow the researcher greater control over the identified set $\mathcal{P}(\Sigma, E)$ and are therefore valuable to her, even though they do not directly enter the decision maker's payoffs. In our lead application to robust causal inference, irrelevance is implied by a mild restriction on the granularity of the assignment mechanism $P(T \mid X)$; in practice, the researcher can *create* payoff-redundant states by conditioning signals on irrelevant variables.

The decision maker's problem The state distribution μ and the kernel of E jointly characterize the decision maker's belief set under information structure (Σ, E). If E is fully informative then the distribution of the state of the world is exactly identified and the decision maker's problem is to choose an action α that maximizes his expected utility with respect to μ. Otherwise, if E is not fully informative then μ is partially identified and the decision maker's problem is to choose α to maximize his worst-case expected utility against the set of priors that are consistent with the data. Formally, given information structure (Σ, E), the *decision maker's problem* is

$$\max_{\alpha \in \Delta(A)} \min_{\nu \in \mathcal{P}(\Sigma, E)} \langle \alpha, \nu \rangle.$$

We emphasize that the decision maker responds to the entire distribution of messages $E\mu$ rather than signal-by-signal to individual messages, and call action α *implementable* if there exists an information structure (Σ, E) under which α is a solution to his problem.

The dual of the decision maker's problem As we show in Lemma 1 in the Appendix, Standard minimax theorems imply that the decision maker's problem of choosing the worst-case optimal action has a *saddle point* (α^*, ν^*) with the property

$$\forall \alpha \in \Delta(A)\ \forall \nu \in \mathcal{P}(\Sigma, E)\ \langle \alpha^*, \nu \rangle \geq \langle \alpha^*, \nu^* \rangle \geq \langle \alpha, \nu^* \rangle.$$

We make extensive use of this fact, and remind the reader that if (α^*, ν^*) is a saddle point then α^* solves the decision maker's problem and ν^* solves the inner minimization problem for α^*. In places, it is convenient to refer to solutions to that problem for the counterfactual environment in which the decision maker believes the state of the world is certainly distributed according to ν. Accordingly, we write

$$\alpha^*(\nu) \equiv \{\alpha \in \Delta(A) | \forall \beta \in \Delta(A)\ \langle \alpha, \nu \rangle \geq \langle \beta, \nu \rangle\}.$$

4 Analysis

The major analytical goals of this paper are to characterize the set of implementable actions and identify the most informative information structures that implement them. Because the identified set $\mathcal{P}(\Sigma, E)$ is completely determined by the fixed prior μ and the kernel of experiment E, the choice of information structure reduces to a choice of kernel.

Feasible kernels satisfy two elementary properties. First, because the map $\nu \mapsto E\nu$ is a linear transformation, $\ker E$ is a linear subspace of $\mathbb{R}^{|\Omega|}$. Second, because every element of $\mathcal{P}(\Sigma, E) \subset \Delta(\Omega)$ is a valid probability distribution and

$$\nu \in \mathcal{P}(\Sigma, E) \iff E\nu = E\mu \iff (\nu - \mu) \in \ker E,$$

it follows that $\sum_\omega d(\omega) = 0$ for every direction $d \in \ker E$. In addition to being necessary, we show in Lemma 3 that the two conditions are jointly sufficient for the existence of an information structure (Σ, E) satisfying $\ker E = D$.

The main result We proceed to our primary characterization Theorem 1. As we show, action α is implementable if and only if there exists a *supporting* prior $\nu \in \mathcal{P}$ satisfying both $\alpha \in \alpha^*(\nu)$ and $\langle \alpha, \mu \rangle \geq \langle \alpha, \nu \rangle$. This condition — which is expressed entirely in terms of primitives — is closely related to Lemma 2, in which we establish that α is a solution to the decision maker's problem under information structure (Σ, E)

if and only if there exists a prior ν in the identified set $\mathcal{P}(\Sigma, E)$ such that (α, ν) form a saddle point.

Theorem 1. *The following three statements are equivalent*

(i) α *is implementable;*

(ii) there exists a prior ν *that supports* α*;*

(iii) there exists a prior ν *that supports* α *and an information structure* (Σ, E) *with* $\ker E = \mathit{span}\{\nu - \mu\}$ *under which* (α, ν) *is a saddle point of the decision maker's problem.*

Theorem 1 does two things. In (ii), we completely characterize implementability via the existence of a supporting prior, as discussed above. Additionally, in (iii), we identify a distinguished class of information structures that implement the desired action while withholding at most one dimension of information from the decision maker. As we establish in our next major result, there is a formal sense in which those information structures are maximally informative among the set

$$I(\alpha) \equiv \{(\Sigma, E) | (\Sigma, E) \text{ implements } \alpha\}$$

of all information structures that implement action α.

Maximally informative implementation We adopt the partial order developed in our earlier paper Rosenthal (2025), in which information structure (Σ, E) is *robustly more informative* than information structure (Σ', E') if $\ker E \subset \ker E'$. We write $(\Sigma, E) \succeq^R (\Sigma', E')$ for the corresponding partial order and equip the set of all information structures with $\succeq^R$. As that paper shows, $(\Sigma, E) \succeq^R (\Sigma', E')$ if and only if for every decision problem (A, u, μ) the worst-case value of that problem under information structure (Σ, E) is greater than or equal to the its worst-case value under information structure (Σ', E'). This order, which is strictly weaker than Blackwell's classical order, is the natural order for our environment.

Proposition 1. *Let action* α *be implementable.*

(i) If μ *supports* α *then* (Σ, E) *is maximally informative for* α *if and only if* (Σ, E) *is fully informative.*

(ii) If μ does not support α then (Σ, E) is maximally informative for α if and only if (Σ, E) is almost fully informative and there exists a supporting prior ν such that $\ker E = $ *span*$\{\nu - \mu\}$.

Much of Proposition 1 follows from Theorem 1, wherein we identify almost fully informative structures that implement each action. The first part is straightforward and the second part follows from the observation that the only proper subspace of $\text{span}\{\nu - \mu\}$ is the trivial kernel $\{0\}$.

The researcher's problem Thus far, we have deliberately focused on the decision maker's problem, for two reasons. First, this emphasis facilitates interpretations in which the information structure (Σ, E) is exogenously specified rather than designed.[1] Second, the researcher's problem can be understood almost immediately through the results we have already developed. As a final matter, we turn our attention to that problem now.

The researcher's utility $v : A \to \mathbb{R}$ is continuous, depends without loss of generality on the decision maker's action only, and her payoff under action α is

$$v(\alpha) \equiv \int_A v(a) \, \mathrm{d}\alpha(a).$$

The *researcher's problem*

$$\max_{(\alpha,\Sigma,E)} \; v(\alpha) \text{ subject to } \forall \beta \in \Delta(A) \; \min_{\nu \in \mathcal{P}(\Sigma,E)} \langle \alpha, \nu \rangle \geq \min_{\nu \in \mathcal{P}(\Sigma,E)} \langle \beta, \nu \rangle$$

is to choose an action α and information structure (Σ, E) that maximizes her payoff $v(\alpha)$ subject to the constraint that (Σ, E) implements α. In light of Theorem 1, this problem is equivalent to maximizing v on the set of implementable actions $\Delta^* \equiv \{\alpha \in \Delta(A) | I(\alpha) \neq \emptyset\}$.

Proposition 2. *There exists an information structure (Σ, E) such that (α, Σ, E) solves the researcher's problem if and only if $\alpha \in \arg\max_{\beta \in \Delta^*} \; v(\beta)$.*

As we show in the proof, standard arguments imply Δ^* is compact. Otherwise, Theorem 1 provides the substance of Proposition 2.

[1] In practice, we imagine a decision maker with access to fixed and potentially incomplete data sets released by a public or private third party.

5 Robust causal inference

Theorem 1 has significant implications for robust causal inference in the canonical treatment effects framework. As we show in Theorem 2, *every* treatment is implementable via an almost fully informative information structure.

Recall the standard causal effects setup. There are finitely many treatments $T_1, ..., T_k$ in set $\mathcal{T}$, finitely many outcomes Y in set $\mathcal{Y} \subset \mathbb{R}$, and finitely many covariates $X \equiv (X_1, ..., X_\ell)$. Each covariate X_j takes values in finite set $\mathcal{X}_j$ and X takes values in $\mathcal{X} \equiv \mathcal{X}_1 \times ... \times \mathcal{X}_\ell$. While the latent state $(Y_1, ..., Y_k, X)$ specifies the potential outcome under each treatment, the observed state $\omega \equiv (Y_T, X, T)$ records only the outcome of the treatment actually received. The researcher does not observe the distribution of the latent state variable but does know the distribution μ of the state variable ω.

The decision maker chooses a treatment $a \in \mathcal{T}$. Common knowledge of both unconfoundedness $(Y_1, ..., Y_k) \perp T \mid X$ and the assignment mechanism (1)

$$P(T = t \mid X = x) \equiv \mu(t|x) \tag{1}$$

with interior support (2)

$$0 < P(T = t|X = x) < 1 \ \ \forall t, x \tag{2}$$

identify each counterfactual mean $E[Y_a]$ from the observed variables (Y_T, X, T). The set of actions A, the state space Ω[2], and the decision maker's utility $u : A \times \Omega \to \mathbb{R}$ are respectively

$$A \equiv \mathcal{T}, \qquad \Omega \equiv \mathcal{Y} \times \mathcal{X} \times \mathcal{T}, \qquad u(a, \omega) \equiv u(a, y, x, t) \equiv \frac{y\mathbf{1}\{a = t\}}{P(T = t \mid X = x)}. \tag{3}$$

For any distribution of observables $\nu \in \Delta(\Omega)$ we have

$$\langle a, \nu \rangle = \mathbb{E}_\nu[u(a, Y, X, T)] = \mathbb{E}_\nu[Y_a],$$

and thus the decision maker's expected utility coincides with the identified causal

[2] We rule out uninteresting cases by assuming $k \geq 2$, $\ell \geq 1$, and that each set $\mathcal{T}, \mathcal{Y}, \mathcal{X}_1, ..., \mathcal{X}_\ell$ contains at least two elements. In turn, the observable state space Ω has at least $2 \times 2 \times 2 = 8$ elements.

effects.

The set of admissible distributions

$$\mathcal{P} \equiv \{\nu \in \Delta(\Omega) \mid \forall t, x\ \nu(t \mid x) = P(T = t|X = x)\} \tag{4}$$

is the set of all distributions ν consistent with the common knowledge assignment mechanism $P(T \mid X)$. In keeping with Assumption 1, we assume that treatment assignment does not depend on at least one covariate. Formally, there exists a variable X_j such that

$$P(T = t \mid X = x) = P(T = t \mid X_{-j} = x_{-j}) \text{ for all } t. \tag{5}$$

If treatment probabilities vary with every covariate, any independent signal satisfies the requirement.

Definition 1. The tuple $(\Omega, A, \mathcal{P}, u, \mu)$ is a *treatment–effects model* if $(\Omega, A, \mathcal{P}, u, \mu)$ satisfies (1)–(5).

As we show in Lemma 5 of the Appendix, treatment-effects models satisfy Assumption 1 and the technical criteria of Section 3. Accordingly, they are well specified special cases of the baseline model.

All actions are implementable The rectangular nature of the state space and the linearity of the decision maker's payoffs impose a great deal of structure on the treatment–effects environment, and the set of payoffs across the set of priors $\mathcal{P}$ is rich. Let $\pi : \mathcal{T} \to \Delta(\mathcal{Y})$ be any family of treatment–outcome marginal distributions, and define the payoff map $U_\pi : \mathcal{T} \to \mathbb{R}$ and the joint distribution ν_π by

$$U_\pi(t) \equiv \sum_y y\pi(y|t), \qquad \nu_\pi(y, x, t) \equiv \pi(y|t)P(T = t \mid X = x)\mu(x).$$

Proposition 3. *Let $(\Omega, A, \mathcal{P}, u, \mu)$ be any treatment–effects model.*

(i) For every family of marginals $\pi : \mathcal{T} \to \Delta(\mathcal{Y})$, the prior ν_π belongs to $\mathcal{P}$ and satisfies $\langle a, \nu_\pi \rangle = U_\pi(a)$ for all $a \in A$.

(ii) For each payoff map $U : \mathcal{T} \to \mathbb{R}$, there exists a family of marginals $\pi : \mathcal{T} \to \Delta(\mathcal{Y})$

such that $U = U_\pi$ if and only if $U(t) \in conv(\mathcal{Y})$ for all $t \in \mathcal{T}$.

Proposition 3 confirms that the priors in $\mathcal{P}$ span the full set of payoff vectors $(\text{conv}(\mathcal{Y}))^{\mathcal{T}}$ consistent with the outcome bounds, independent of the structural restrictions the assignment mechanism $P(T \mid X)$ imposes on $\mathcal{P}$. In conjunction with Theorem 1, this implies that every action in a treatment–effects model is implementable.

Theorem 2. *Let $(\Omega, A, \mathcal{P}, u, \mu)$ be any treatment–effects model. Every action α in $\Delta(A)$ is implementable.*

Our approach to proving Theorem 2 is constructive. First, we construct a payoff vector U satisfying (i) $U(\alpha) \geq U(\beta)$ for all $\beta \in \Delta(A)$ and (ii) $U(\alpha) \leq \langle \alpha, \mu \rangle$. Second, we apply Proposition 3 to extract a prior ν_π that induces U. In turn, because ν_π supports α, Theorem 1 provides an almost fully informative information structure (Σ, E) with $\ker E = \text{span}\{\nu_\pi - \mu\}$ that implements α.

Marginal information structures We apply our formal results to the natural class of *marginal information structures* that disclose the joint distribution of a non-empty strict subset of the observable variables. In order to state the definition, for each variable $v \in \{Y, X_1, ..., X_\ell, T\}$ we write $\mathcal{X}_v$ for its domain, where $\mathcal{X}_{X_j} \equiv \mathcal{X}_j$, $\mathcal{X}_Y \equiv \mathcal{Y}$, and $\mathcal{X}_T \equiv \mathcal{T}$.

Definition 2. Let $V \subset \{Y, X_1, ..., X_\ell, T\}$ be a strict and non-empty subset of the observable variables and write

$$\Omega^V \equiv \prod_{v \in V} \mathcal{X}_v$$

for the coarsened state space Ω^V. The *marginal information structure* (Σ^V, E^V) associated with V satisfies $\Sigma^V \equiv \Omega^V$ and $E^V(v|\omega) \equiv \mathbf{1}\{\omega_V = v\}$ for each state ω. For each distribution $\nu \in \Delta(\Omega)$, the distribution of messages generated by the marginal information structure (Σ^V, E^V) discloses the marginal distribution

$$(E^V \mu)(v) = \sum_{\omega : \omega_V = v} \mu(\omega)$$

of the variables in V.

Proposition 4. *Consider any treatment–effects model $(\Omega, A, \mathcal{P}, u, \mu)$ and let V be any non-empty strict subset of $\{Y, X_1, ..., X_\ell, T\}$. The marginal information structure*

(Σ^V, E^V) *is not maximally informative for implementing any action* α.

Proposition 4 implies that marginal disclosure imposes unnecessary uncertainty on the decision maker. We illustrate this point by returning to the motivating example and constructing an almost fully informative information structure preserves the implementation of the researcher's preferred intervention while weakly improving the decision maker's welfare.

Example Return to Section 2. As demonstrated, the marginal information structure $(\Sigma^{\{Y,T\}}, E^{\{Y,T\}})$ implements the researcher's preferred treatment $T = 1$ while full disclosure does not. At the same time, $(\Sigma^{\{Y,T\}}, E^{\{Y,T\}})$ is not almost fully informative, and thus Proposition 4 implies there exists a strictly more informative structure that also implements $T = 1$ while preserving or improving the decision maker's payoff. In particular, any information structure (Σ, E) satisfying

$$E(\cdot \mid 0,0,0) + E(\cdot \mid 1,1,0) = E(\cdot \mid 1,0,0) + E(\cdot \mid 0,1,0)$$

implements $T = 1$. These information structures conceal only the statistic

$$\Pr(Y = 1 \mid X = 1, T = 0) - \Pr(Y = 1 | X = 0, T = 0) \tag{6}$$

summarizing baseline outcome heterogeneity between the two groups. Compare this to $(\Sigma^{\{Y,T\}}, E^{\{Y,T\}})$, under which there is uncertainty about (6) *and* group composition

$$Pr(X = 0), \qquad\qquad Pr(X = 1).$$

The former is sufficient to embed the supporting prior for $a = 1$ into the belief set $\mathcal{P}(\Sigma, E)$, and concealing the distribution of covariates serves no purpose other than to potentially degrade the decision maker's payoff for each of the actions he might take.

6 Discussion and conclusions

This paper studies a prior-free model of information design in which the decision maker's belief set is the set of all priors consistent with the observed distribution of messages generated by an information structure (Σ, E). If the associated transition matrix has full rank, then the state distribution ν is exactly identified and the decision problem reduces to standard expected-utility maximization. Otherwise, ν is partially identified and the decision maker evaluates actions according to their worst-case payoff over the identified set.

We interpret the information structure as a choice variable, characterize the set of implementable actions, and identify the most informative structures that implement them. The condition for implementability is straightforward to verify, and the maximally informative experiments that implement them have a closed form.

Our framework applies directly to causal inference. We formalize this connection, demonstrate via example that there are interesting policy reversals under partial disclosure, and show in broad generality that all treatments are implementable via an appropriately designed information structure. The maximally informative information structure for implementing each action is strictly more informative than disclosure of the marginal distribution of any proper subset of the variables in the data set. In developing the application, we offer a unified perspective on strategic data disclosure and robust causal inference.

References

Bergemann, D. and S. Morris (2016). Bayes correlated equilibrium and the comparison of information structures. *Theoretical Economics 11*(4), 1461–1513.

Bergemann, D. and S. Morris (2019). Information design: A unified perspective. *Journal of Economic Literature 57*(1), 44–95.

Blackwell, D. (1951). Comparison of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 93–102.

Blackwell, D. (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics 24*, 265–272.

Çelen, B. (2012). Informativeness of experiments for MEU. *Journal of Mathematical Economics 48*(6), 404–406.

Heyen, D. and B. R. Wiesenfarth (2015). Informativeness of experiments for MEU – a recursive definition. *Journal of Mathematical Economics 57*, 28–30.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101 101*(6), 2590–2615.

Kline, B. and E. Tamer (2023). Recent developments in partial identification. *Annual Review of Economics 15*, 125–150.

Li, J. and J. Zhou (2016). Blackwell's informativeness ranking with uncertainty-averse preferences. *Games and Economic Behavior 96*, 18–29.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review 80*(2), 319–323.

Manski, C. F. (1997). Monotone treatment response. *Econometrica 65*(6), 1311–1334.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. New York: Springer-Verlag.

Manski, C. F. (2007). *Identification for Prediction and Decision*. Cambridge, MA: Harvard University Press.

Manski, C. F. (2013). *Public Policy in an Uncertain World*. Harvard University Press.

Molinari, F. (2020). Microeconometrics with partial identification. In D. L. Millimet, B. H. Baltagi, and I. A. Prucha (Eds.), *Handbook of Econometrics, Volume 7B*, pp. 355–486. North-Holland.

Rosenthal, M. (2025). Blackwell without priors. Manuscript, arXiv:2510.08709 [v2: Mon, 13 Oct 2025 11:40:58 UTC].

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics 2*, 167–195.

Whitmeyer, M. (2025). Blackwell-monotone updating rules. *Journal of Political Economy*. Forthcoming.

Yang, K. H., N. Yoder, and A. Zentefis (2025). Explaining models. Technical report, SSRN. Working Paper, revised May 18, 2025; SSRN:4723587; accepted to ACM EC 2025.

A Appendix

A.1 Omitted lemmas

Lemma 1. *The decision maker's problem under information structure* (Σ, E) *has a saddle point.*

Proof. First, the set of actions $\Delta(A)$ is compact and convex because the set of primitive actions A is a compact metric space. Second, because $\mathcal{P}$ is compact and convex and the map $\nu \mapsto E\nu$ is linear, the identified set $\mathcal{P}(\Sigma, E)$ is compact and convex. Third, the map $\alpha \mapsto \langle \alpha, \nu \rangle$ is linear (and hence both continuous and concave) for each ν and the map $\nu \mapsto \langle \alpha, \nu \rangle$ is linear (and hence both continuous and convex) for each α. Sion's minimax theorem implies the decision maker's problem has a saddle point (α^*, ν^*). $\square$

Lemma 2. *If* α *solves the decision maker's problem under information structure* (Σ, E) *then there exists* $\nu \in \mathcal{P}(\Sigma, E)$ *such that* (α, ν) *is a saddle point of the decision maker's problem.*

Proof. Let α solve the agent's problem; let (α^*, ν^*) be a saddle point in that problem as provided by Lemma 1; let ν be a minimizer for $\langle \alpha, \cdot \rangle$ in $\mathcal{P}(\Sigma, E)$; and note

$$\langle \alpha, \nu \rangle = \max_{\alpha' \in \Delta(A)} \min_{\nu' \in \mathcal{P}(\Sigma, E)} \langle \alpha', \nu' \rangle = \min_{\nu' \in \mathcal{P}(\Sigma, E)} \max_{\alpha' \in \Delta(A)} \langle \alpha', \nu' \rangle = \langle \alpha^*, \nu^* \rangle$$

because α is a solution and (α^*, ν^*) is a saddle point. First, because ν is a minimizer for $\langle \alpha, \cdot \rangle$; $\langle \alpha, \nu \rangle = \langle \alpha^*, \nu^* \rangle$; and α^* is a maximizer for $\langle \cdot, \nu^* \rangle$, we have

$$\forall \beta \in \Delta(A) \; \langle \alpha, \nu^* \rangle \geq \langle \alpha, \nu \rangle = \langle \alpha^*, \nu^* \rangle \geq \langle \beta, \nu^* \rangle.$$

Second, because α^* is a maximizer for $\langle \cdot, \nu^* \rangle$; $\langle \alpha, \nu \rangle = \langle \alpha^*, \nu^* \rangle$; and ν is a minimizer for $\langle \alpha, \cdot \rangle$, we have

$$\forall \gamma \in \mathcal{P}(\Sigma, E) \; \langle \alpha, \nu^* \rangle \leq \langle \alpha^*, \nu^* \rangle = \langle \alpha, \nu \rangle = \langle \alpha, \gamma \rangle.$$

Accordingly, (α, ν^*) is a saddle point, as claimed. $\square$

Lemma 3. *There exists an information structure* (Σ, E) *with* $\ker E = D$ *if and only if (i)* D *is a linear subspace of* $\mathbb{R}^{|\Omega|}$ *and (ii)* $\sum_{\omega\in\Omega} d(\omega) = 0$ *for all directions* $d \in D$.

Proof. Let (Σ, E) be any information structure and note $\ker E$ is a linear subspace of $\mathbb{R}^{|\Omega|}$ because E is linear. For each $d \in \ker E$, we have

$$\sum_{\omega\in\Omega} d(\omega) = \sum_{\omega\in\Omega} d(\omega) \sum_{\sigma\in\Sigma} E(\sigma|\omega) = \sum_{\sigma\in\Sigma} (Ed)(\sigma) = \sum_{\sigma\in\Sigma} 0 = 0,$$

where the first equality follows from $\sum_\sigma E(\sigma|\omega) = 1$ for each state ω and the third from $d \in \ker E$.

Conversely, let D satisfy (i), (ii) and note (ii) implies $\dim(D) \leq |\Omega| - 1$ because D is a subset of the $|\Omega| - 1$ dimensional hyperplane $H \equiv \{d \in \mathbb{R}^{|\Omega|} | \sum_\omega d(\omega) = 0\}$. There are two trivial cases. First, if $\dim(D) = |\Omega| - 1$ then $D = H$ and thus any information structure with $\Sigma \equiv \{\sigma\}$ and $E(\sigma|\omega) \equiv 1$ for all states ω satisfies $\ker E = D$. Second, if $\dim(D) = 0$ then $D = \{0\}$ and any fully informative information structure (Σ, E) satisfies $\ker E = D$.

Suppose instead $0 < \dim(D) < |\Omega| - 1$ and write $n \equiv |\Omega|$. We proceed constructively. Let W be the orthogonal complement of D, note $\ell \equiv \dim(W) = n - \dim(D)$, and let $w^1, ..., w^\ell$ be a basis for W. For each $1 \leq i \leq \ell$, let scalars x^i, y^i satisfy $x^i > -\min(w^i)$ and $y^i > \max(w^i)$. Define

$$p^i \equiv (x^i, ..., x^i) + w^i, \qquad q^i \equiv (y^i, ..., y^i), \qquad \lambda \equiv \frac{1}{\sum_i (x^i + y^i)}.$$

Let M be the $m \times n$ real matrix with $m \equiv 2\ell$ rows each defined by

$$(M_{i1}, ..., M_{in}) \equiv \begin{cases} \lambda p^i & i \leq \ell \\ \lambda q^{i-\ell} & i > \ell. \end{cases}$$

Finally, let Σ be any set of messages with m distinct elements, enumerate Ω by $\omega_1, ..., \omega_n$, and define experiment E by $E(\cdot|\omega_j) \equiv (M_{1j}, ..., M_{mj})$ for each column j of M. Because the entries of M are non-negative per our choice of p^i, q^i and

$$\sum_{i=1}^{m} M_{ij} = \sum_{i=1}^{l} \lambda(x_i + y_i) = 1,$$

(Σ, E) is a valid information structure. We claim $\ker E = D$. First, because $w^1, ..., w^\ell$ is a basis for W and

$$w^i = \frac{y^i p^i - x^i q^i}{x^i + y^i}$$

for each i, the set W is contained in the row space of M. Second, because $(1, ..., 1) \in W$ per our hypothesis $\sum_\omega d(\omega) = 0$ for all $d \in D$ and $w^i \in W$ for all i by definition, we have $(M_{i1}, ..., M_{in}) \in W$ for all i. Accordingly, the row space of M is contained in W. Finally, because we have just established that the row space of M is the orthogonal complement of D, the fundamental theorem of linear algebra implies $\ker E = D$. $\square$

Lemma 4. *For each prior $\nu \in \mathcal{P}$ there exists a prior $\nu' \in \mathcal{P}$ such that (i) $\langle \alpha, \nu' \rangle = \langle \alpha, \nu \rangle$ for all actions α and (ii) $\mu + \lambda(\nu' - \mu) \in \mathcal{P}$ only if $\lambda \leq 1$.*

Proof. Let state ω satisfy $\mu(\omega) > 0$. There are two cases. First, if $\nu(\omega) = 0$ then set $\nu' \equiv \nu$ and note

$$(\mu + \lambda(\nu' - \mu))(\omega) = (1 - \lambda)\mu(\omega) + \lambda\nu'(\omega) = (1 - \lambda)\mu(\omega).$$

Accordingly, if $\lambda > 1$ then $(\mu + \lambda(\nu' - \mu))(\omega) < 0$ and hence $(\mu + \lambda(\nu' - \mu)) \notin \mathcal{P} \subset \Delta(\Omega)$. Criterion (ii) follows; criterion (i) is vacuous for $\nu' = \nu$. Second, if $\nu(\omega) > 0$ then the first part of Assumption 1 implies there exists a state ω' with $u(a, \omega') = u(a, \omega)$ for all a. Define

$$\nu'(s) \equiv \begin{cases} 0 & s = \omega \\ \nu(\omega) + \nu(\omega') & s = \omega' \\ \nu(s) & s \neq \omega, \omega'. \end{cases}$$

and note (i) $\langle \alpha, \nu' \rangle = \langle \alpha, \nu \rangle$ for all actions α and (ii) the second part of Assumption 1 implies $\nu' \in \mathcal{P}$. The claim reduces to the first case. $\square$

Lemma 5. *Let $(\Omega, A, \mathcal{P}, u, \mu)$ be any treatment–effects model*

(i) $(\Omega, A, \mathcal{P}, u, \mu)$ satisfies the technical criteria of Section 3 and

(ii) $(\Omega, A, \mathcal{P}, u, \mu)$ satisfies Assumption 1.

Proof. The sets $\Omega, A, \mathcal{P}$ are non-empty; Ω and A are finite; and utility u is continuous.

The belief set

$$\mathcal{P} = \{\nu \in \Delta(\Omega) | \nu(t|x) = P(T = t \mid X = x)\}$$

is the intersection of the compact and convex simplex $\Delta(\Omega)$ with a finite set of linear equality constraints delineated by the assignment mechanism $P(T \mid X)$ and is therefore itself compact and convex. This establishes (i). Turning to (ii), hypothesis (5) implies the existence of a covariate X_j such that

$$x_{-j} = x'_{-j} \implies P(T = t|X = x) = P(T = t|X = x') \ \forall t.$$

Accordingly, for any two states $\omega = (y, x, t)$ and $\omega' = (y, x', t)$ that satisfy $x_{-j} = x'_{-j}$, the definition (3) of u implies $u(a, \omega) = u(a, \omega')$ for all actions $a \in A$ and therefore confirms the existence of payoff-equivalent states. In order to verify the second part of Assumption 1, recall the constraints that define $\mathcal{P}$ depend only on the conditional treatment problems $P(T \mid X)$, which (5) implies are constant across states ω, ω' that are identical in components y, t, and x_{-j}. Accordingly, if $\nu, \nu' \in \Delta(\Omega)$ differ only in their allocation of probability mass within such pairs of states, then $\nu \in \mathcal{P} \iff \nu' \in \mathcal{P}$. We confirm that the treatment effects model satisfies the second part of Assumption (1). □

A.2 Omitted proofs of propositions and theorems

Proof of Theorem 1. We show (i) $\implies$ (ii) $\implies$ (iii) $\implies$ (i). To see the first implication, suppose information structure (Σ, E) implements α. Lemma 1 and Lemma 2 imply there exists a prior $\nu \in \mathcal{P}(\Sigma, E)$ such that (α, ν) are a saddle point of the decision maker's problem. Because α maximizes $\langle \cdot, \nu \rangle$ on $\Delta(A)$ and ν minimizes $\langle \alpha, \cdot \rangle$ on $\mathcal{P}(\Sigma, E)$, it follows immediately that ν supports α. We conclude (ii) holds.

To see the second implication, suppose ν supports α. By Lemma 4 there exists a prior $\nu' \in \mathcal{P}$ such that (i) $\langle \alpha, \nu' \rangle = \langle \alpha, \nu \rangle$ for all actions α and (ii) $\mu + \lambda(\nu' - \mu) \in \mathcal{P}(\Sigma, E) \implies \lambda \leq 1$. To ease notation, rewrite $\nu \equiv \nu'$ and apply Lemma 3 to obtain an information structure (Σ, E) with $\ker E = \operatorname{span}\{\nu - \mu\}$. We claim ν minimizes $\langle \alpha, \cdot \rangle$ on $\mathcal{P}(\Sigma, E)$. Towards that end, choose $\nu' \in \mathcal{P}(\Sigma, E)$ arbitrarily and note $\ker E = \operatorname{span}\{\nu - \mu\}$ implies there exists a scalar λ such that $\nu' = \mu + \lambda(\nu - \mu)$, where taking expectations yields

$$\langle \alpha, \nu' \rangle = \langle \alpha, \mu \rangle + \lambda(\langle \alpha, \nu \rangle - \langle \alpha, \mu \rangle).$$

First, because ν supports α, we have $\langle \alpha, \nu \rangle - \langle \alpha, \mu \rangle \leq 0$. In turn, $\lambda \leq 1$ implies

$$\langle \alpha, \mu \rangle + \lambda(\langle \alpha, \nu \rangle - \langle \alpha, \mu \rangle) \geq \langle \alpha, \nu \rangle.$$

Accordingly, (α, ν) are a saddle point of the decision maker's problem under information structure (Σ, E) and (iii) holds. The third implication (iii) implies (i) is immediate. $\square$

Proof of Proposition 1. Suppose α is implementable. We make implicit use of Lemma 3 throughout. First, if $\alpha \in \alpha^*(\mu)$ then (i) any information structure (Σ, E) with $\ker E = \{0\}$ implements α and (ii) $\ker E \subset \ker E'$ for all information structures (Σ', E'). Conversely, if (Σ, E) is an element of $I(\alpha)$ and $\dim(\ker E) \geq 1$ then any information structure (Σ', E') with $\ker E' = \{0\}$ belongs to $I(\alpha)$ and satisfies $(\Sigma', E') \succ^R (\Sigma, E)$.

Second, if $\alpha \notin \alpha^*(\mu)$ then let ν support α. The only information structures (Σ, E) for which $\ker E$ is a proper subset of $\operatorname{span}\{\nu - \mu\}$ are those with $\ker E = \{0\}$. Because such information structures do not implement α per our hypothesis $\alpha \notin \alpha^*(\mu)$, we conclude that (Σ, E) is a maximal element of $I(\alpha)$ if $\ker E = \operatorname{span}\{\nu - \mu\}$. Conversely, if (Σ, E) implements α and $\ker E$ is not of the asserted form for any prior ν that supports α, Lemma 2 implies there exists a prior $\nu' \in \mathcal{P}(\Sigma, E)$ such that (α, ν') is a saddle point of the decision maker's problem under that information structure. Because it follows immediately that ν' supports α, Theorem 1 provides an information structure (Σ', E') that implements α with $\ker E' = \operatorname{span}\{\nu' - \mu\}$. Because $\nu' \in \mathcal{P}(\Sigma, E)$, we have $\operatorname{span}\{\nu' - \mu\} \subset \ker E$; because (Σ, E) does not satisfy $\ker E = \operatorname{span}\{\nu' - \mu\}$ by hypothesis, the inclusion is strict. In turn, because $(\Sigma', E') \in I(\alpha)$ and $(\Sigma', E') \succ^R (\Sigma, E)$, the given information structure (Σ, E) is not a maximal element of $I(\alpha)$. $\square$

Proof of Proposition 2. First, Theorem 1 implies that there exists an information structure (Σ, E) implementing α if and only if $\alpha \in \Delta^*$. Second, Δ^* is the projection of the set $G \equiv \{(\alpha, \nu) \in \Delta(A) \times \mathcal{P} | \alpha \in \alpha^*(\nu), \langle \alpha, \mu \rangle \geq \langle \alpha, \nu \rangle\}$ onto its first coordinate. The correspondence α^* has a closed graph by the maximum theorem; the map

$(\alpha, \nu) \mapsto \langle \alpha, \mu \rangle - \langle \alpha, \nu \rangle$ is continuous; G is therefore a closed subset of a compact set and thus itself compact; and hence Δ^* is the projection of a compact set and therefore compact. Because Δ^* is compact and v is continuous by hypothesis, the researcher's problem has a solution. □

Proof of Proposition 3. We proceed with (i). First, because $\nu_\pi(y, x, t) \geq 0$ for all y, x, t and

$$\sum_{y,x,t} \nu_\pi(y, x, t) = \sum_{y,x,t} \pi(y|t) P(t \mid x) \mu(x) = \sum_{x,t} P(t \mid x) \mu(x) = \sum_x \mu(x) = 1,$$

we have $\nu_\pi \in \Delta(\Omega)$. Second, because

$$\nu_\pi(t|x) = \frac{\sum_y \nu_\pi(y, x, t)}{\sum_\tau \sum_y \nu_\pi(y, x, \tau)} = \frac{\sum_y \pi(y|t) P(t \mid x) \mu(x)}{\sum_\tau P(\tau \mid x) \mu(x)} = \frac{P(t \mid x) \mu(x)}{\mu(x)} = P(t \mid x),$$

we have $\nu_\pi \in \mathcal{P}$. Third,

$$\langle a, \nu_\pi \rangle = \sum_{y,x,t} \frac{y \mathbf{1}\{a = t\} \nu_\pi(y, x, t)}{P(t \mid x)} = \sum_{y,x} \frac{y \pi(y|a) P(a \mid x) \mu(x)}{P(a \mid x)} = \sum_x U_\pi(a) \mu(x) = U_\pi(a).$$

Accordingly, ν_π induces payoff $\langle a, \nu_\pi \rangle = U_\pi(a)$ for each $a \in A$. Part (ii) follows immediately from the definition of $\mathrm{conv}(\mathcal{Y})$. □

Proof of Theorem 2. Choose action $\alpha \in \Delta(A)$ and write

$$u^* \equiv \min_{t \in \mathrm{supp}(\alpha)} \mathbb{E}_\mu[Y_t]$$

for the worst counterfactual outcome in its support. Let $U^* : \mathcal{T} \to \mathrm{conv}(\mathcal{Y})$ be any payoff map satisfying (i) $U^*(t) = u^*$ for all $t \in \mathrm{supp}(\alpha)$ and (ii) $U^*(t) \leq u^*$ for all $t \notin \mathrm{supp}(\alpha)$. First, Proposition 3 provides a family of marginals π and a prior $\nu_\pi \in \mathcal{P}$ with $\langle a, \nu_\pi \rangle = U_\pi(a)$ for all $a \in A$. Second, because

$$\langle \alpha, \nu_\pi \rangle = \sum_{a \in A} U^*(a) \alpha(a) \geq \sum_{a \in A} U^*(a) \beta(a) = \langle \beta, \nu_\pi \rangle$$

for all actions $\beta \in \Delta(A)$, we have $\alpha \in \alpha^*(\nu_\pi)$. Third, our choice of π implies

$$\langle \alpha, \nu_\pi \rangle = \sum_{a \in A} U_\pi(a)\alpha(a) \leq \sum_{a \in A} \mathbb{E}_\mu[Y_a]\alpha(a) = \sum_{a \in A} \langle a, \mu \rangle \alpha(a) = \langle \alpha, \mu \rangle.$$

We conclude that ν_π supports α. Apply Theorem 1 to obtain an information structure (Σ, E) with $\ker E = \text{span}\{\nu_\pi - \mu\}$ that implements α. $\square$

Proof of Proposition 4. We claim $\dim(\ker E) > 1$; the result then follows immediately Proposition 1 because maximally informative information structures have $\dim(\ker E) \leq 1$. Towards that end, let v be any variable excluded from V and let k_v be the number of elements in its domain, recalling that V is a strict subset of $\{Y, X_1, ..., X_\ell, T\}$ by hypothesis. Because $E^V(\cdot|\omega) = E_j^V(\cdot|\omega')$ for every pair of states ω, ω' that differ only in component v, and because there are $|\Omega|/k_v$ such equivalence classes of states, the image of Ω under experiment E^V contains at most $|\Omega|/k_v$ distinct message distributions. Accordingly,

$$\dim(\ker E^V) \geq \frac{k_v - 1}{k_v}|\Omega| \geq \frac{1}{2}|\Omega| \geq 4,$$

where the first inequality is by the rank-nullity theorem, the second by $k_v \geq 2$ for all variables v, and the third by $|\Omega| \geq 8$. We conclude that (Σ^V, E^V) is not maximally informative. $\square$