Private Private Information

Kevin He Fedor Sandomirskiy Omer Tamuz

April 28, 2022
Virtual Seminars in Economic Theory (VSET)

A Familiar Setup: Private Information about State

- A finite set of agents $\{1, ..., n\}$
- A **binary state** of nature $\omega \in \{\ell, h\}$
- A common prior probability $p \in (0, 1)$ for event $\{\omega = h\}$
- Agent i gets signal $s_i \in S_i$ about ω, her **private information**
- A joint distribution $\mathbb{P}$ over $(\omega, s_1, ..., s_n)$ defines the **information structure**

Private Information May Not Be Private

- Some examples of private information structures:
  - **Public** signals — $\mathbb{P}[s_1 = s_2 = ... = s_n] = 1$
  - **Conditionally independent** signals — given ω, $(s_1, ..., s_n)$ are drawn independently across agents
- Are agents' information in these examples really private?
- Clearly, public signals are not private at all
- Even conditionally independent signals are not very private
  - Suppose prior $\mathbb{P}[\omega = h] = 1/2$
  - Binary signals with $\mathbb{P}[s_i = \omega \mid \omega] = 3/4$
  - Before observing s_1, P1 assigns belief $1/2$ to $\{s_2 = h\}$
  - After learning $s_1 = h$, P1 now assigns belief $5/8$ to $\{s_2 = h\}$
  - s_1 contains info about s_2, so P2's info not fully private after all

Private Private Information

Definition

A ***private*** **private information structure** is one where the signals $(s_1, ..., s_n)$ are independent.

- Signals must be independent, not conditionally independent
- Private private signals contain info about the state, but not about each other
- Signals do not update agents' higher-order posterior beliefs: i learns nothing about j's belief from s_i
- Is it possible for everyone to have **informative** private private signals?
  - May seem paradoxical at first: s_1 informative about ω, ω correlated with s_2, yet P1 learns nothing about s_2?
  - It is possible!
  - Tension between informativeness and privacy: impossible for everyone to have **perfectly** informative private private signals
  - We focus on this tension and study how informative private private signals can be

Application of **Informative** Private Private Signals

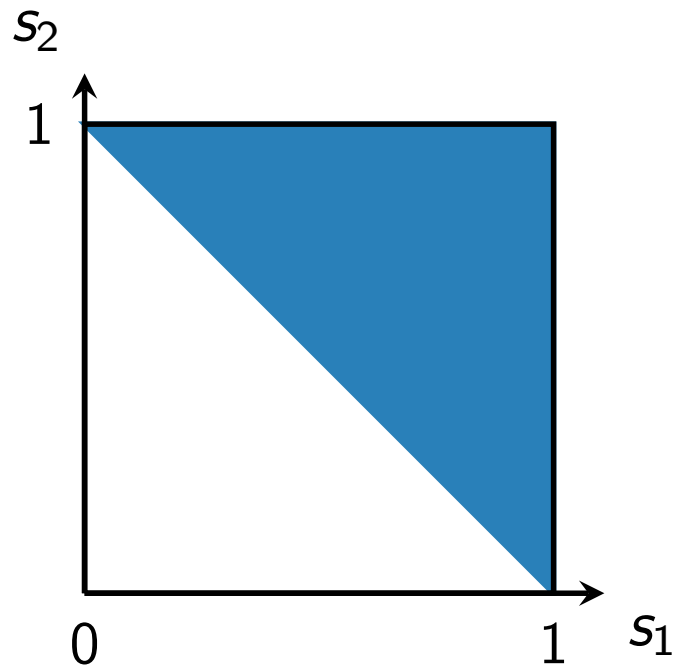
Fairness, equity, and privacy in rating design:

- State $\omega \in \{\ell, h\}$ is borrower's creditworthiness
- s_1 is a private or legally protected trait, correlated with ω
- Rating agency knows ω and s_1, bank knows neither
- Rating agency gives bank a signal s_2 about the borrower
- Regulations / privacy laws may require s_2 to be independent of s_1 (guarantees a fairness concept called **demographic parity**)
- So, (ω, s_1, s_2) is a private private info structure
- How informative can s_2 be?

Outline

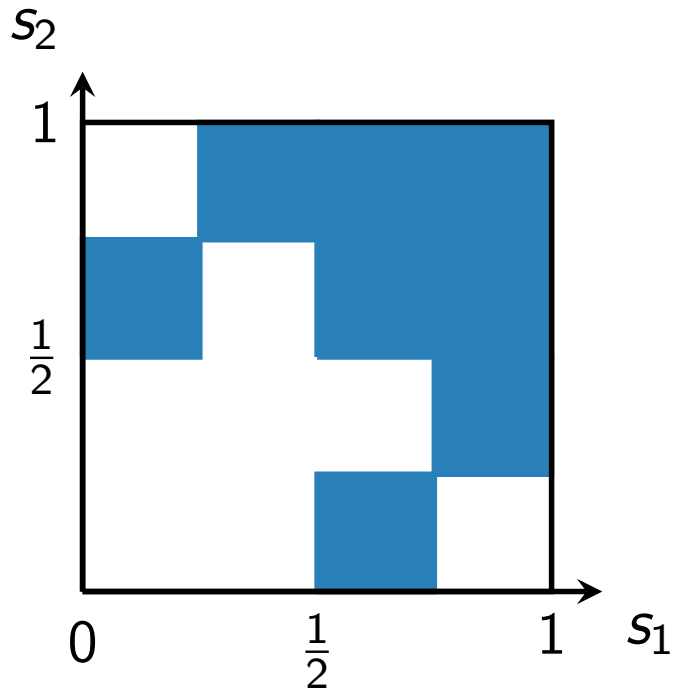
1. Canonical representation of private private info structures
2. Pareto optimal private private info structures
3. Application: most informative signal under the constraint of not revealing a protected trait

Example 1



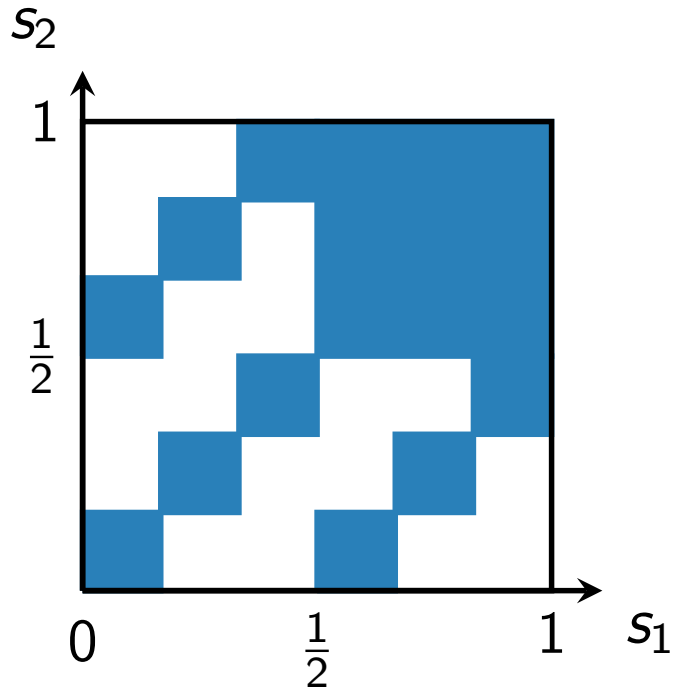
- Suppose $\mathbb{P}[\omega = h] = 1/2$
- When $\omega = h$, choose (s_1, s_2) uniformly from blue set
- When $\omega = \ell$, choose (s_1, s_2) uniformly from white set
- s_1, s_2 independent, each $s_i \sim \text{Unif}[0, 1]$
- $\{\omega = h\}$ is the event that $s_1 + s_2 \geq 1$
- $\mathbb{P}[\omega = h \mid s_1 = 0.9]$ fraction of vertical slice at 0.9 that is blue
- In fact, each s_i induces posterior belief s_i about state

Example 2



- Same idea: $\mathbb{P}[\omega = h] = 1/2$, draw (s_1, s_2) uniformly from blue or white set depending on state
- $s_i \overset{\text{i.i.d.}}{\sim} \text{Unif}[0,1]$, but s_i no longer induces belief s_i
- Equivalent to a binary signal that induces beliefs $1/4$ or $3/4$

Example 3



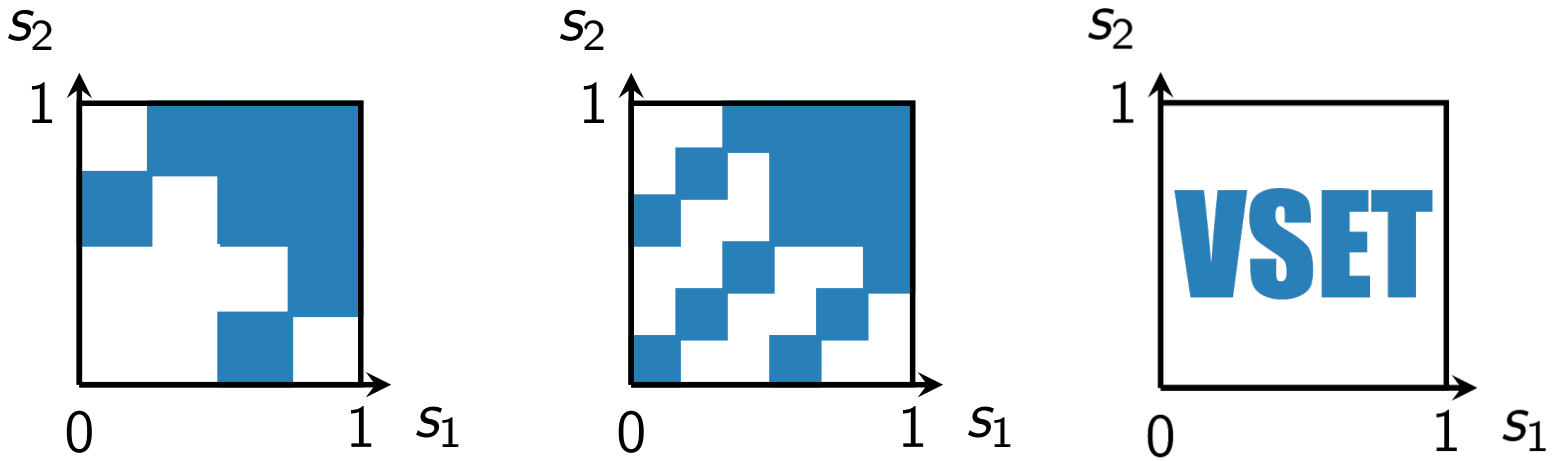
- And here is an example that induces beliefs 1/3 or 2/3

Canonical Representation of Private Private Signals

A general method to construct private private info structures:

- Suppose $\mathbb{P}[\omega = h] = p$ and there are n agents
- Fix any subset of $[0,1]^n$ with measure p, call it A
- When $\omega = h$, choose $(s_1, ..., s_n)$ uniformly from A
- When $\omega = \ell$, choose $(s_1, ..., s_n)$ uniformly from $[0,1]^n \backslash A$
- (Equivalently, $s_i \overset{\text{i.i.d.}}{\sim} \text{Unif}[0,1]$, ω given by whether $\vec{s} \in A$)

Call this the **canonical** info structure associated with A — as we vary A, get different private private info structures



- Is this just a large family of examples, or is this "everything"?

Canonical Representation of Private Private Signals

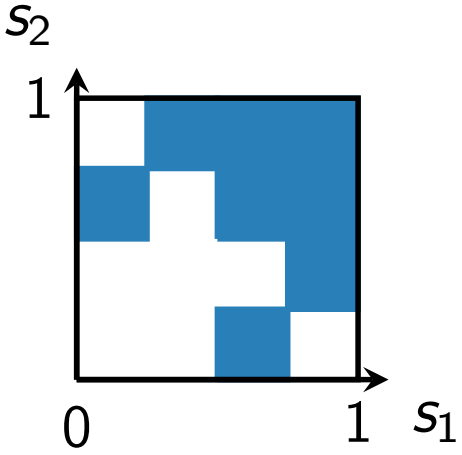
Proposition

For every private private info structure, there is an equivalent canonical info structure — a measurable $A \subseteq [0,1]^n$ that induces the same belief distribution for each agent as the given structure.

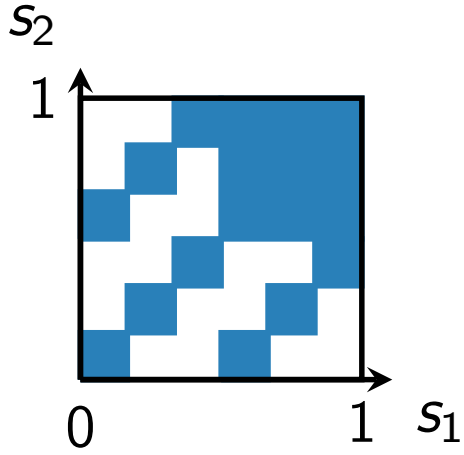
Proof idea (for $n = 2$):

- Can always assume $s_i \sim \text{Unif}[0,1]$ by relabeling signals
- Key restriction: state determined by signal realizations (s_1, s_2)
- Given any private private info structure (ω, s_1, s_2), let $f(s_1, s_2) := \mathbb{P}[\omega = h \mid s_1, s_2]$
- Consider (ω, s_1, s_2, s_3) where $s_3 \sim \text{Unif}[0,1]$, $\omega = h$ iff $s_3 \leq f(s_1, s_2)$, so s_3 **resolves uncertainty** left by (s_1, s_2)
- Construct a canonical private private info structure by splitting the realization of s_3 among P1 and P2 via **secret sharing**

Pareto Comparisons of Informativeness



Ex. 2: Belief **1/4** or **3/4**



Ex. 3: Belief **1/3** or **2/3**

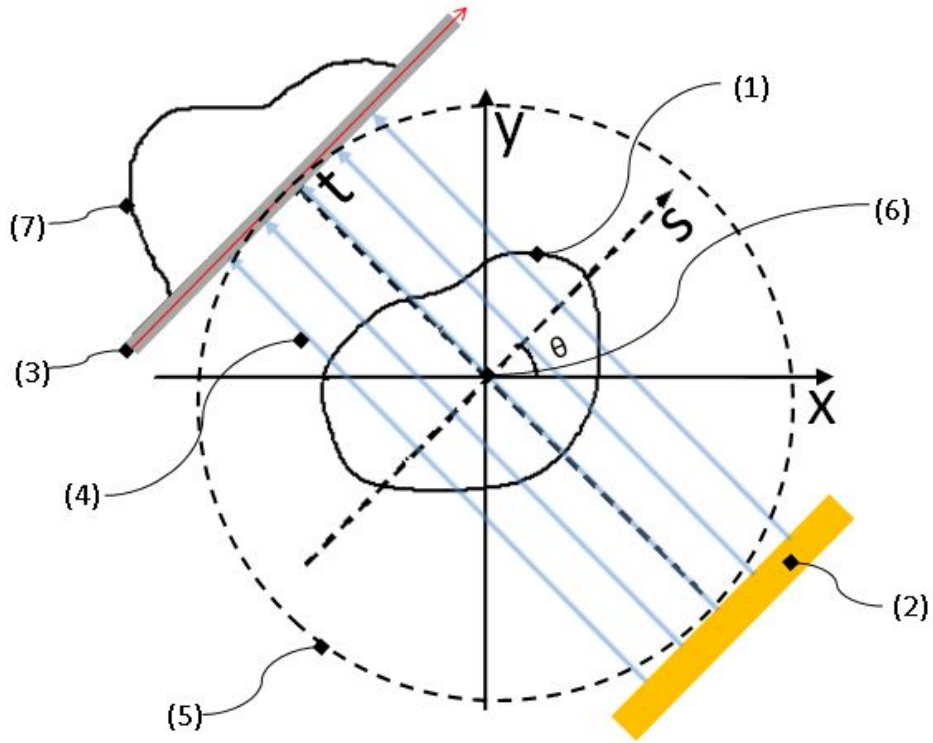
- Each agent's info in Example 2 strictly Blackwell dominates her info in Example 3
- Suppose we want to give out as much info as possible. What are the **most informative** private private info structures?

Definition

For I, I' private private info structures, $I \succeq I'$ if each agent's info about state in I Blackwell dominates her info about state in I'. A private private info structure is **Pareto optimal** if it is $\succeq$-maximal (cannot give any agent more info without violating privacy).

Tomography and Sets of Uniqueness

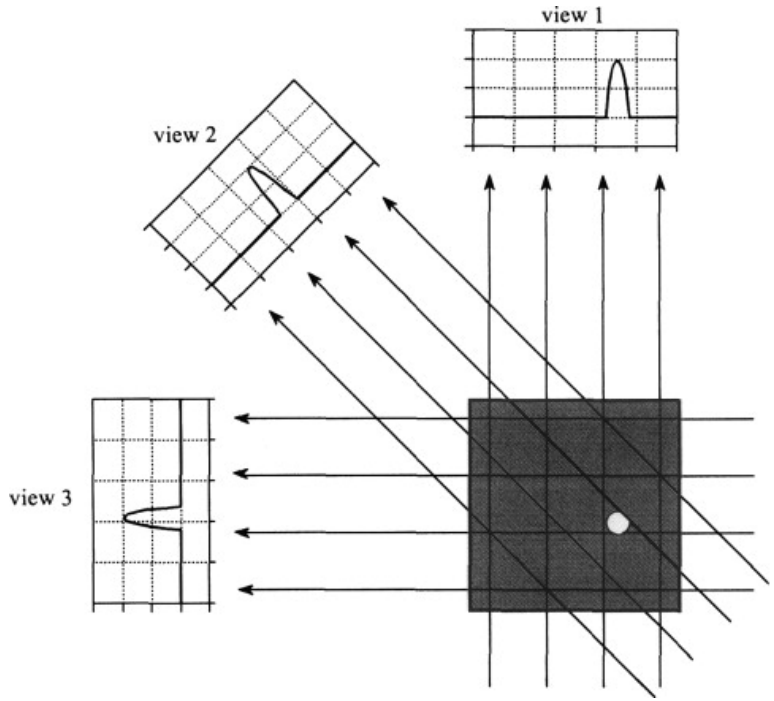
- **Tomography** is an imaging technique that investigates the shape of an object by running x-ray through it



- Produces a lower-dimensional projection of the object by looking at how much x-ray is absorbed at different points

Tomography and Sets of Uniqueness

- Typically, must run x-ray from many different angles to get a good understanding of the object's geometry



Definition

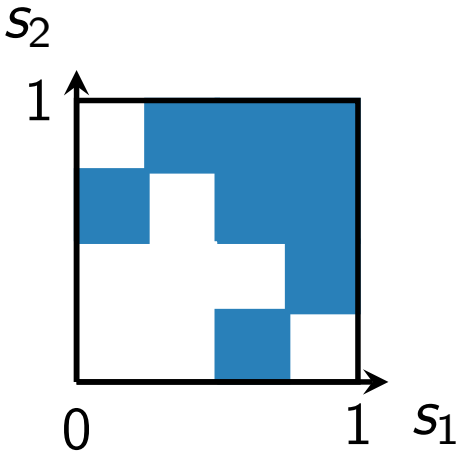
$A \subseteq [0,1]^n$ is a **set of uniqueness** if its n projections onto the n coordinate axes suffice to reconstruct A. (That is, if A' matches A on all n coordinate-axes projections, then $A' = A$ a.e. in $[0,1]^n$.)

Pareto Optimality and Sets of Uniqueness

Theorem

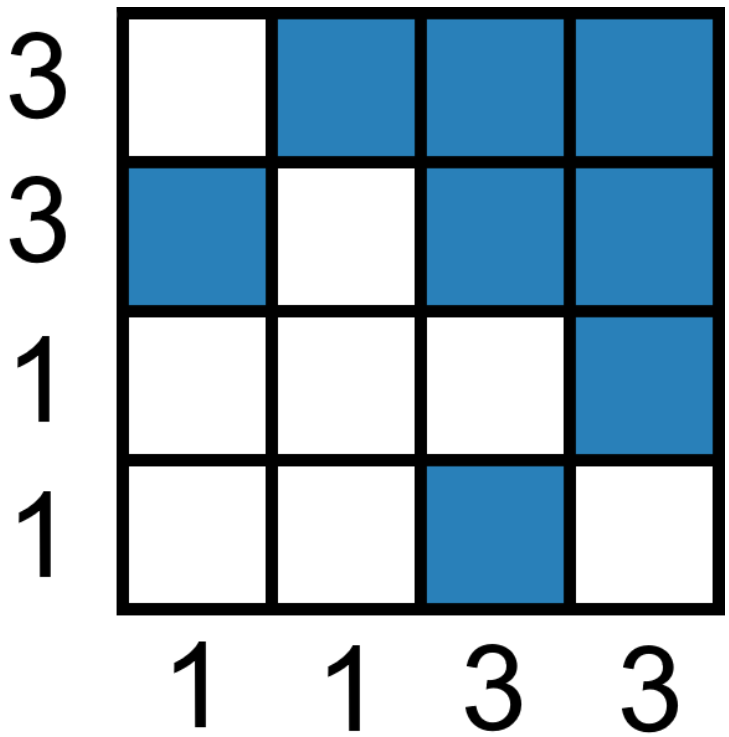
The private private info structure associated with $A \subseteq [0,1]^n$ is Pareto optimal if and only if A is a set of uniqueness.

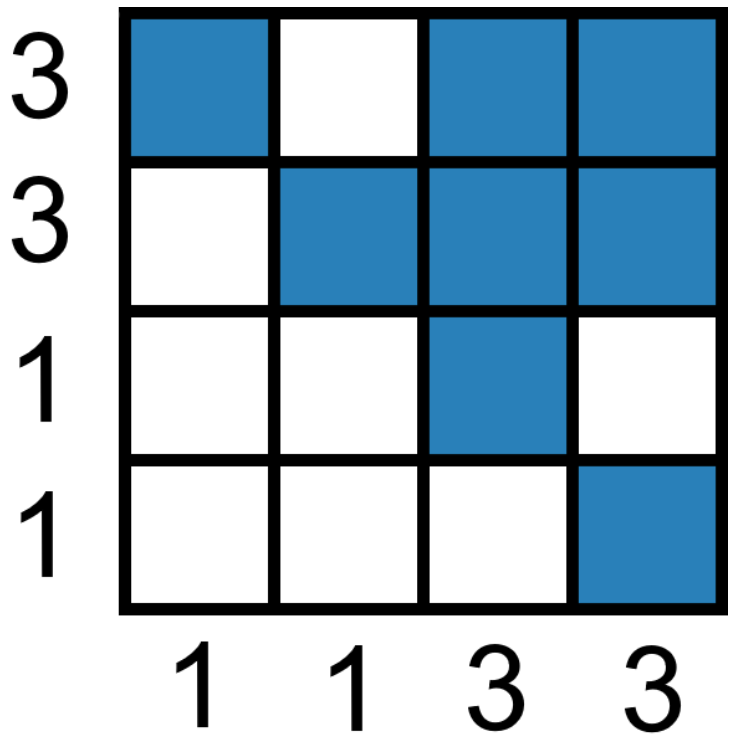
- An unexpected connection between Pareto optimality of private private info structures and a concept from tomography
- Will discuss its proof later (if there's time at the end)
- As an application, recall Example 2 strictly Pareto dominates Example 3. But is Example 2 itself Pareto optimal?
- Question equivalent to: is the blue area a set of uniqueness?



A Puzzle!

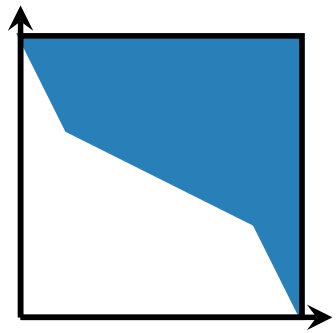
Problem for a newspaper puzzle column: is there another coloring of the 4x4 grid that preserves all column-wise and row-wise counts of colored squares?





Existing Results about Sets of Uniqueness

- By our theorem, this shows the binary info structure that induces beliefs $1/4$ or $3/4$ is not itself Pareto optimal
- Can **disprove** Pareto optimality by finding another set with same marginal projections. How to **prove** Pareto optimality?
- Use results about sets of uniqueness from tomography
- $A \subseteq [0,1]^n$ is **upward closed** if $\vec{x} \in A \Rightarrow \vec{x}' \in A$ for all $\vec{x}' \geq \vec{x}$



- $A \subseteq [0,1]^n$ is **additive** if there are bounded, non-decreasing $h_i : [0,1] \to \mathbb{R}$ s.t.

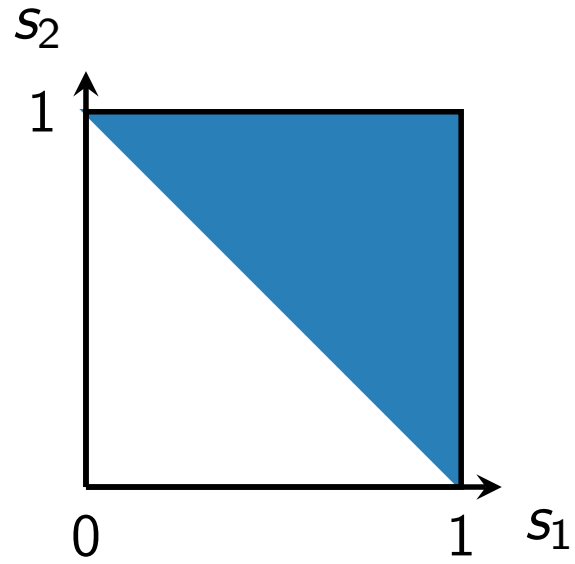
$$A = \{\vec{x} \in [0,1]^n : \sum_{i=1}^{n} h_i(x_i) \geq 0\}$$

- Additive implies upward closed, equivalent if $n = 2$

Existing Results about Sets of Uniqueness

Theorem (Fishburn, Lagarias, Reeds, and Shepp (1990))

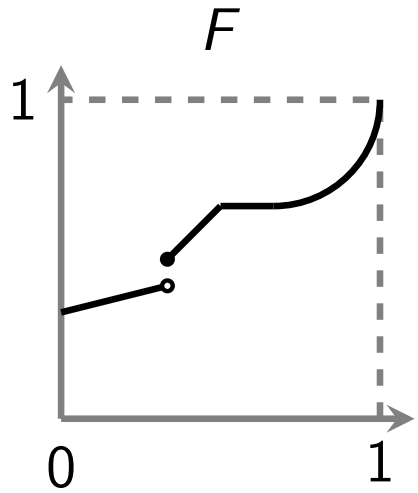
- *For $n = 2$, every upward closed set is a set of uniqueness, and every set of uniqueness is upward closed up to measure-preserving transformations of axes.*
- *For every n, every additive set is a set of uniqueness.*

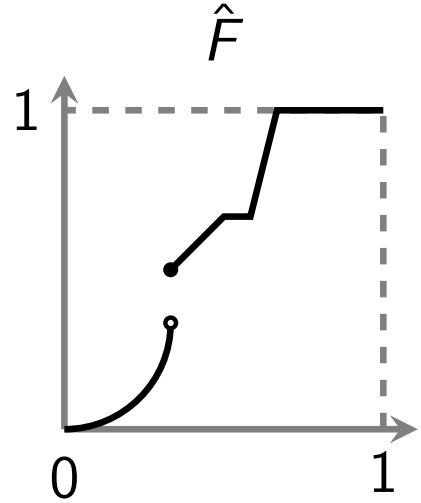


- Blue set in Example 1 is upward closed, so its info structure is Pareto optimal (apply Fishburn et al.'s theorem, then ours)
- For any number of agents n, can use additive subsets of $[0,1]^n$ to generate Pareto optimal private private signals
- For two agents, upward closed subsets give **all** possible belief distributions in Pareto optimal private private info structures

Conjugates and Pareto Optimality for $n = 2$

- As a corollary, get simple test of Pareto optimality when $n = 2$
- Let F be the cdf of a distribution on $[0, 1]$ with mean p
- Denote $\hat{F}(x) = 1 - F^{-1}(1 - x)$
- Then $\hat{F}$ is also the cdf of a distribution on $[0, 1]$ with mean p, obtained by reflecting F around the anti-diagonal





- Call F and $\hat{F}$ **conjugates**

Corollary

For $n = 2$, a private private info structure is Pareto optimal if and only if the belief distributions induced by s_1 and s_2 are conjugates.

Application: Optimal Signal Given a Protected Trait

- ω — a binary state of interest
- s_1 — a sensitive or protected trait
- Fix the joint distribution of (ω, s_1)
- Want to generate a signal s_2 that is
  - **as informative as possible** about ω
  - but **independent** of s_1
- Equivalently: find a **Pareto optimal** private private info structure $\mathbb{P}$ on (ω, s_1, s_2) with the **given** (ω, s_1) **marginal**

Proposition

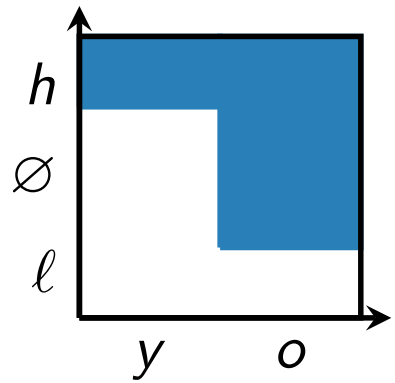
There is a Pareto optimal private private info structure on (ω, s_1, s_2) with the given (ω, s_1) marginal, and it is unique up to equivalence. Belief distributions induced by s_1 and s_2 are conjugates in this structure.

Informativeness of Trait and Optimal Signal

- $\omega \in \{\ell, h\}$ fit for job, $s_1 \in \{y, o\}$ binary measure of age
- $\mathbb{P}[\omega = h] = \mathbb{P}[s_1 = o] = 1/2$
- Age is correlated with fit

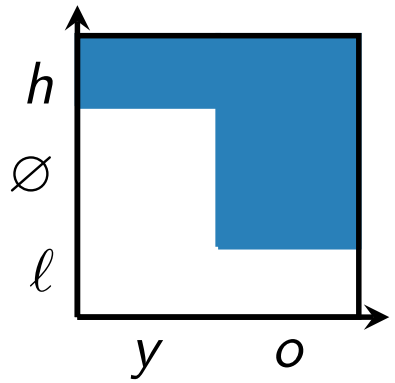
$$\mathbb{P}[\omega = h \mid s_1 = o] = 1/2 + r = \mathbb{P}[\omega = \ell \mid s_1 = y]$$

  for some $0 \leq r \leq 1/2$
- Here is the optimal signal that preserves privacy (why?)



- It has the correct marginal distribution on (ω, s_1):
  - $\mathbb{P}[\omega = h] = 1/2$ (half of the square is blue)
  - $\mathbb{P}[\omega = h \mid s_1 = o] = \mathbb{P}[\omega = \ell \mid s_1 = y] = 1/2 + r$
- Blue area is upward closed $\Rightarrow$ associated with a Pareto optimal private private info structure (Fishburn et al. + our theorem)

Informativeness of Trait and Optimal Signal



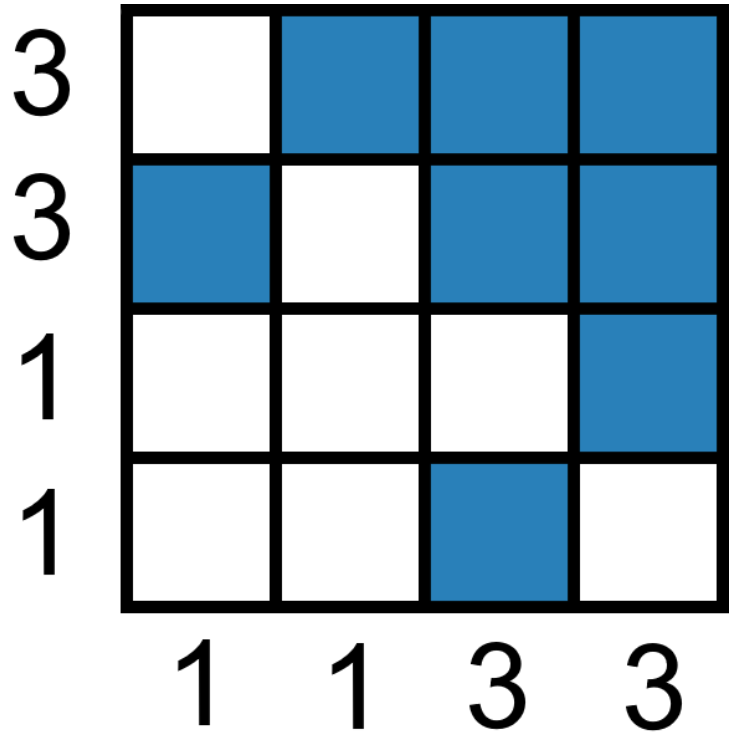
- Read off the solution from this picture
- Most informative signal that does not reveal age is **trinary**
  - $S_2 = \{\ell, \varnothing, h\}$
  - $s_2 = \varnothing$ is uninformative about fit
  - $s_2 \in \{\ell, h\}$ perfectly reveals the fit
  - $\mathbb{P}[s_2 = \varnothing] = 2r$: if age more correlated with fit, then less info can be sent about job fit without violating privacy

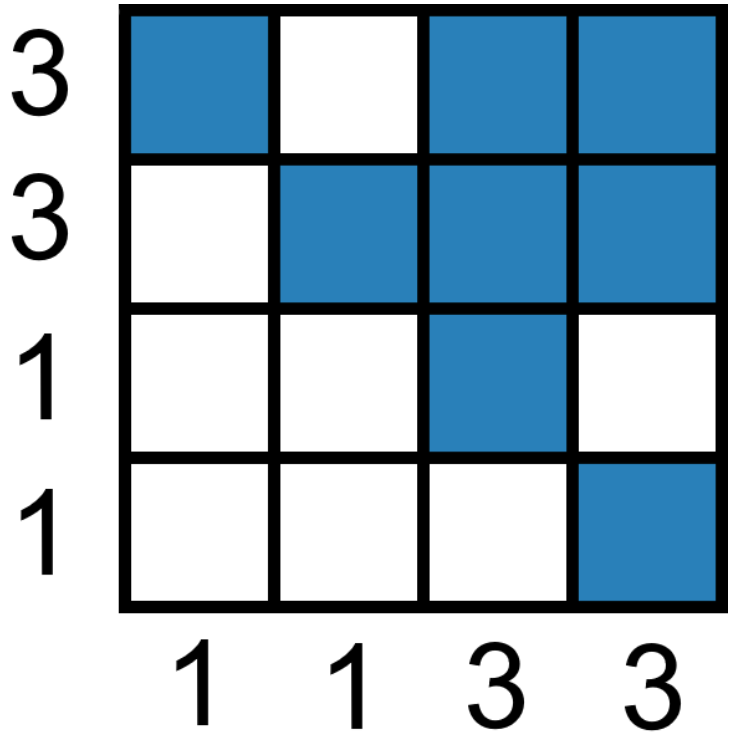
Connecting Pareto Optimality with Tomography

Theorem

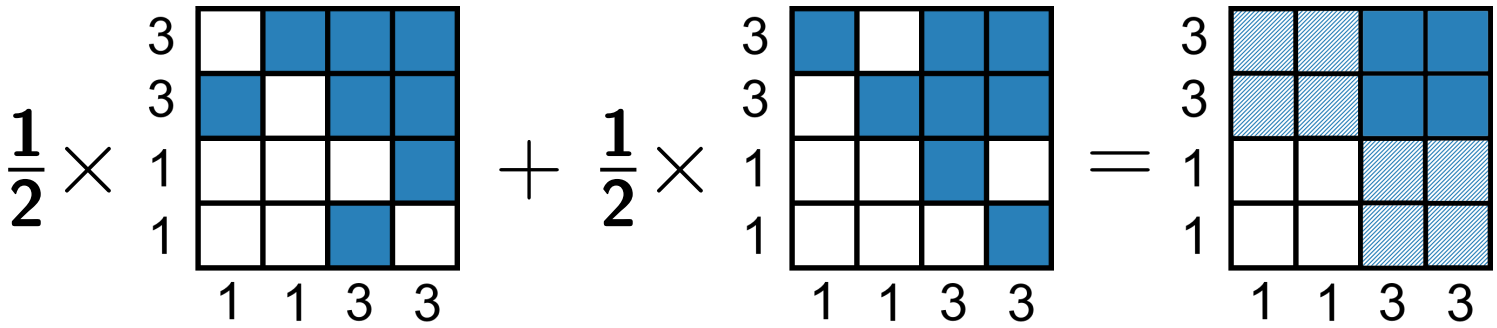
The private private info structure associated with $A \subseteq [0,1]^n$ is Pareto optimal if and only if A is a set of uniqueness.

Key idea: can use two sets A, A' with same marginals to build a strictly Pareto dominating private private info structure





The "Convex Combination" Coloring



- Each square can now be colored, blank, or **shaded**
- Shaded square = "half of a colored square"
- Draw $s_1, s_2 \overset{\text{i.i.d.}}{\sim} \text{Unif}[0,1]$. If (s_1, s_2) in shaded region, toss an independent fair coin s_3 to determine state ω
- This structure generates the same distribution of posteriors
  - Because "convex combination" between the two colorings with the same marginals preserves the marginals

Converse: Set of Uniqueness $\Rightarrow$ Pareto Optimal

- $A \subseteq [0,1]^n$ is a set of uniqueness
- $I =$ associated private private info structure, $s_i \in S_i = [0,1]$
- By way of contradiction, suppose some private private info structure I^{dom} with signals T_i strictly Pareto dominates I
- By Blackwell's theorem, find garblings $\varphi_i : T_i \to \Delta([0,1])$ s.t. $\varphi_i(t_i)$ gives the same posterior distribution as s_i
- Consider the info structure I^{garb} where $(t_1, ..., t_n)$ generated as in I^{dom}, then agents observe $(\varphi_1(t_1), ..., \varphi_n(t_n))$
- I^{garb} is private private and equivalent to I, and WLOG can reparametrize signals so that signals in I^{garb} are uniform on $[0,1]$ and each signal x_i gives the same posterior belief as in I
- Since at least one φ_i **strictly** garbles, for positive measure of $\vec{x}$, $\mathbb{P}[\omega = h \mid (\varphi_1(t_1), ..., \varphi_n(t_n)) = (x_1, ..., x_n)] \notin \{0,1\}$

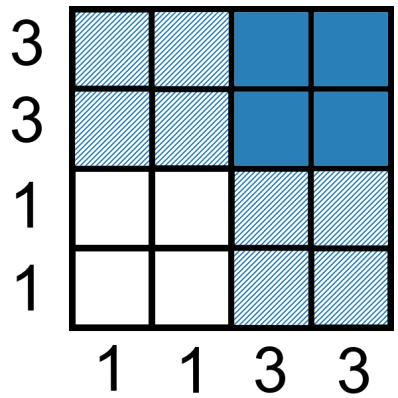
Converse: Set of Uniqueness $\Rightarrow$ Pareto Optimal

- Define $f : [0,1]^n \to [0,1]$, with

$$f(x_1, ..., x_n) := \mathbb{P}[\omega = h \mid (\varphi_1(t_1), ..., \varphi_n(t_n)) = (x_1, ..., x_n)]$$

- Note f has the same projections on coordinate axes as 1_A
  - Almost contradicts A being a set of uniqueness, but f is not indicator on a set
- Gutmann, Kemperman, Reeds, and Shepp (1991): Let $\mathbb{F}$ = set of functions $[0,1]^n \to [0,1]$ whose coordinate axes projections agree with 1_A. **Indicator functions = extreme points** of $\mathbb{F}$
- Since $f \in \mathbb{F}$ is not an extreme point $\Rightarrow$ $\mathbb{F}$ has non-empty relative interior $\Rightarrow$ there are at least 2 distinct extreme points $\mathbf{1}_{V'}, \mathbf{1}_{V''} \in \mathbb{F}$
- At least one of V' or V'' is a set with the same marginals as A but does not equal to A a.e., contradiction

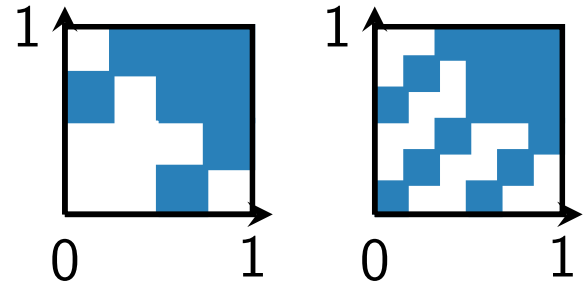
Not a Set of Uniqueness $\Rightarrow$ Strictly Dominated

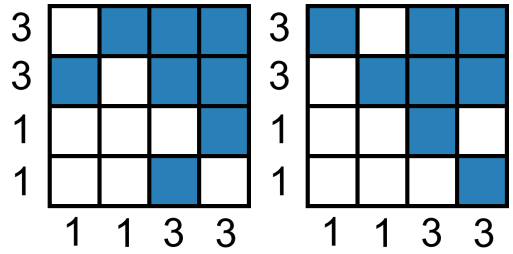


- Toss the fair coin ahead of time and tell P2 how it lands
- Let $s_1' = s_1$, $s_2' = (s_2, s_3)$
- Signals of P1 and P2 still independent, because the coin is independent of s_1, s_2 — a private private info structure
- (s_2, s_3) strictly Blackwell dominates s_2 because the coin affects the state with positive probability (shaded region)
  - Uses the hypothesis that A is not a set of uniqueness

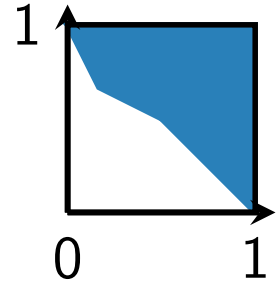
Summary

- ***Private* private information structures**: signals of different agents $(s_1, s_2, ..., s_n)$ are (unconditionally) independent
- Can **represent** all such info structures as subsets of $[0, 1]^n$ (up to equivalence)
- Pareto optimal private private info structures associated with **sets of uniqueness**: subsets that are determined by their projections on coordinate axes
- For $n = 2$, set A associated with Pareto optimal private private info structure iff A is **upward closed** (up to relabeling)
  - So, given a pre-existing signal s_1, most informative s_2 that is independent of s_1 induces its conjugate belief distribution
- For any n, **additive** sets associated with Pareto optimal private private structures





(not Pareto optimal)



(Pareto optimal)