Unintended Consequences

Causality, Misperception, & Games

Joe Halpern
Cornell University
Computer Science

Evan Piermont
Royal Holloway, University of London
Department of Economics

Marie-Louise Vierø
Aarhus University
Department of Economics

PPE Society London --- July 2025

ATTEMPTING *to dissuade pregnant women from consuming alcohol while pregnant, many US states have passed laws classifying drinking during pregnancy as child abuse/neglect. Such laws are robustly associated with an increase in adverse birth outcomes, i.e., low birth-weight or preterm births.*[1]

[1] Subbaraman, Meenakshi S., and Sarah CM Roberts. "Costs associated with policies regarding alcohol use during pregnancy: results from 1972-2015 Vital Statistics." PloS one 14.5 (2019).

DURING *Company rule of colonial India, the Governors-General became concerned of the proliferation of cobras. Embracing some newfangled capitalism, he came up with what appeared an ingenious solution — pay a bounty to any man who brought forward a head of a dead cobra. This "cash for snakes" program worked for a while, but was followed by a marked uptick in the number of cobras plaguing the city.*[2]

[2] Source: probably fiction

SEEING *that the average sparrow eats 2kg of grain per year, the Maoist regime instituted the “eliminate sparrows campaign” encouraging the citizenry to hit noisy pots and pans so as to prevent sparrows from resting in their nests, with the goal of causing them to drop dead from exhaustion. After pushing sparrow populations to near extinction within China, crop yields plummeted.*[3]

[3] Source: copy/paste from wikipedia

In each vignette, a decision maker misunderstands the causal structure of the environment:

- Alcoholic women, afraid of prosecution, stopped going to the prenatal doctors appointments
- Locals, enticed by the bounty, started breeding cobras
- Sparrows eat locusts; locusts bad

In this paper we

- model games (i.e., strategic environments) as causal structures
- allow misperception the causal structure
- examine what kind of misperceptions are sustainable

A (non-strategic) **causal structure** is

- a directed acyclic graph of variables $\mathcal{V}$
  - Say $Y \lhd X$ if edge from Y to X
  - Let $\mathcal{R}(X)$ collect the set of values $X \in \mathcal{V}$ can take
- a structural equation for each variable

$$F_X : \prod_{\substack{Y \in \mathcal{V} \\ Y \lhd X}} \mathcal{R}(Y) \to \mathcal{R}(X)$$

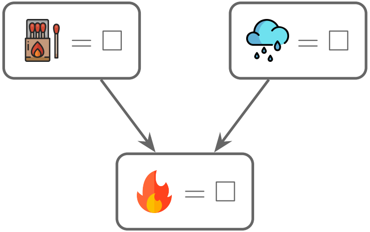
Example

◇ Three variables: $X_{🌧}$, $X_{🧯}$, $X_{🔥}$

$$X_{🧯} = 1 \qquad (F_{X_{🧯}})$$

$$X_{🌧} = 0 \qquad (F_{X_{🌧}})$$

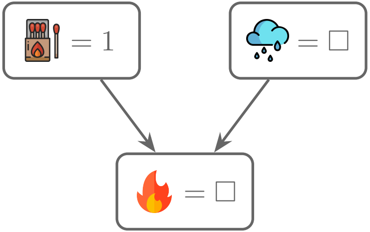
$$X_{🔥} = \min\{X_{🧯}, 1 - X_{🌧}\} \qquad (F_{X_{🔥}})$$



Example

- Three variables: $X_{\text{🌧}}, X_{\text{🧯}}, X_{\text{🔥}}$

$$
\begin{aligned}
X_{\text{🧯}} &= 1 && (F_{X_{\text{🧯}}})\\
X_{\text{🌧}} &= 0 && (F_{X_{\text{🌧}}})\\
X_{\text{🔥}} &= \min\{\ 1\ \ , 1 - X_{\text{🌧}}\} && (F_{X_{\text{🔥}}})
\end{aligned}
$$



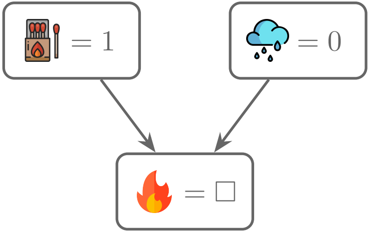
Example

- Three variables: $X_{🌧}$, X_{matches}, $X_{🔥}$

$$X_{\text{matches}} = 1 \qquad (F_{X_{\text{matches}}})$$

$$X_{🌧} = 0 \qquad (F_{X_{🌧}})$$

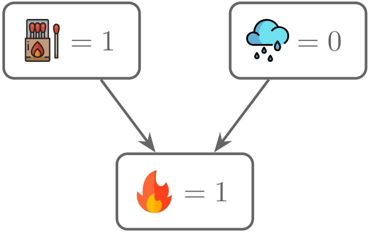
$$X_{🔥} = \min\{\ 1\ \ , 1 -\ \ 0\ \ \} \qquad (F_{X_{🔥}})$$



Example

- Three variables: $X_{\text{🌧️}}, X_{\text{match}}, X_{\text{🔥}}$

$$
\begin{aligned}
X_{\text{match}} &= 1 && (F_{X_{\text{match}}}) \\
X_{\text{🌧️}} &= 0 && (F_{X_{\text{🌧️}}}) \\
X_{\text{🔥}} &= 1 && (F_{X_{\text{🔥}}})
\end{aligned}
$$



A **causal game** is

- A set of players, $\mathcal{I}$
- A partially ordered set variables $(\mathcal{V}, \lhd\,)$ (with specified ranges)
- Utility function for each player:

$$u_i : \prod_{X \in \mathcal{V}} \mathcal{R}(X) \to \mathbb{R}$$

In addition, players are assigned a set of variables which they **control**:

- $A_i \subseteq \mathcal{V}$ are variables controlled by $i \in \mathcal{I}$
- $A_i \cap A_j = \varnothing$
- The variables $\mathcal{V} \setminus \bigcup_{i \in \mathcal{I}} A_i$ are uncontrolled
  - Represent nature's choices / structural components of the game
  - Each uncontrolled $X \in \mathcal{V}$ is dictated by a structural equation F_X

Example

There are two player's: the politician (p) and the mother (m) and 5 variables:

- X: health of the infant — range $= \{-1, 0, 1\}$
- X: mother drinks while pregnant — range $= \{0, 1\}$
- X: law making it illegal to drink while pregnant — range $= \{0, 1\}$
- X: mother goes to the doctor — range $= \{0, 1\}$
- X: mother is punished for breaking the law — range $= \{0, 1\}$

Example

The utility functions are:

$$u_{\texttt{p}}(\vec{X}) = X_{\text{💓}},$$

and

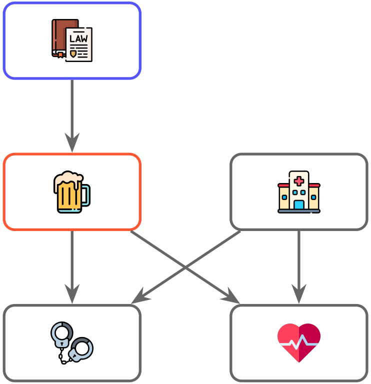
$$u_{\texttt{m}}(\vec{X}) = (3 \times (1 - X_{\text{⛓}})) + (2 \times X_{\text{🍺}}) + X_{\text{💓}}$$

Example

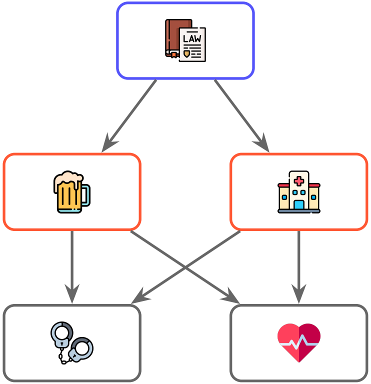
			$u_{\texttt{m}}$	$u_{\texttt{p}}$
0	0	-1	2	-1
0	0	0	3	0
0	0	1	4	1
0	1	-1	4	-1
0	1	0	5	0
0	1	1	6	1
1	0	-1	-1	-1
1	0	0	0	0
1	0	1	1	1
1	1	-1	1	-1
1	1	0	2	0
1	1	1	3	1

- A **strategy** for i is a choice is structural equation for each $X \in A_i$
  - must respect $\lhd$
- An **equilibrium** is a profile of strategies such that no player can change her structural equations to improve her payoff.

Example



Exogenous Doc Choice: $\lhd_{exo}$



Endogenous Doc Choice: $\lhd_{endo}$

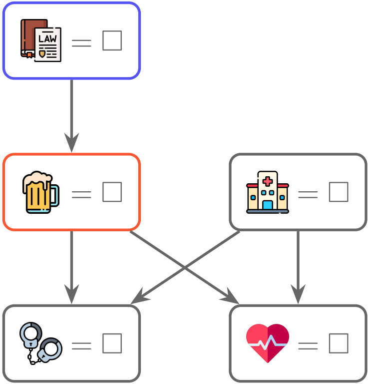
Example: Exogenous Choice

- Uncontrolled Variables: $X_{\text{hospital}}, X_{\text{arrest}}, X_{\text{health}}$

$$X_{\text{hospital}} = 1 \qquad (F_{X_{\text{hospital}}})$$

$$X_{\text{arrest}} = \min\{X_{\text{hospital}}, X_{\text{beer}}, X_{\text{law}}\} \qquad (F_{X_{\text{arrest}}})$$

$$X_{\text{health}} = X_{\text{hospital}} - X_{\text{beer}}, \qquad (F_{X_{\text{health}}})$$



Example: Exogenous Choice

- Uncontrolled Variables: $X_{\text{hospital}}, X_{\text{handcuffs}}, X_{\text{heart}}$

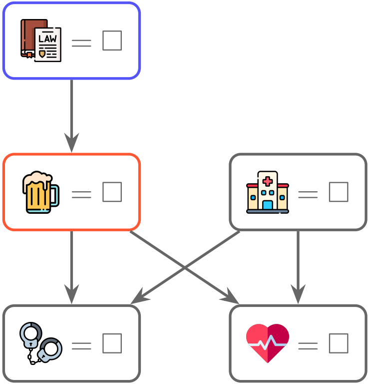
$$X_{\text{hospital}} = 1 \qquad (F_{X_{\text{hospital}}})$$

$$X_{\text{handcuffs}} = \min\{X_{\text{hospital}}, X_{\text{beer}}, X_{\text{law}}\} \qquad (F_{X_{\text{handcuffs}}})$$

$$X_{\text{heart}} = X_{\text{hospital}} - X_{\text{beer}}, \qquad (F_{X_{\text{heart}}})$$

- Strategy for `m`

$$X_{\text{beer}} = 1 - X_{\text{law}} \qquad (F_{X_{\text{beer}}})$$



Example: Exogenous Choice

- Uncontrolled Variables: $X_{🏥}, X_{⛓}, X_{❤}$

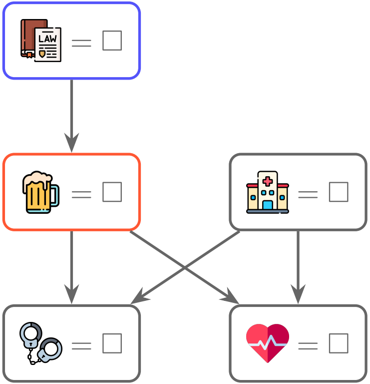
$$X_{🏥} = 1 \qquad (F_{X_{🏥}})$$
$$X_{⛓} = \min\{X_{🏥}, X_{🍺}, X_{📕}\} \qquad (F_{X_{⛓}})$$
$$X_{❤} = X_{🏥} - X_{🍺}, \qquad (F_{X_{❤}})$$

- Strategy for $\texttt{m}$

$$X_{🍺} = 1 - X_{📕} \qquad (F_{X_{🍺}})$$

- Strategy for $\texttt{p}$

$$X_{📕} = 1 \qquad (F_{X_{📕}})$$



Example: Exogenous Choice

- Uncontrolled Variables: $X_{🏥}, X_{🚔}, X_{❤️}$

$$X_{🏥} = 1 \qquad (F_{X_{🏥}})$$

$$X_{🚔} = \min\{\ 1\ \ , X_{🍺},\ \ 1\ \ \} \qquad (F_{X_{🚔}})$$

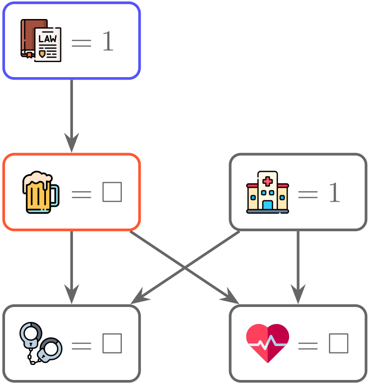
$$X_{❤️} = \ \ 1\ \ - X_{🍺}, \qquad (F_{X_{❤️}})$$

- Strategy for m

$$X_{🍺} = 1 - 1 \qquad (F_{X_{🍺}})$$

- Strategy for p

$$X_{📕} = 1 \qquad (F_{X_{📕}})$$



Example: Exogenous Choice

- Uncontrolled Variables: X, X, X

$$X = 1 \qquad (F_X)$$

$$X = \min\{\ 1\ ,\ 0\ ,\ 1\ \} \qquad (F_X)$$

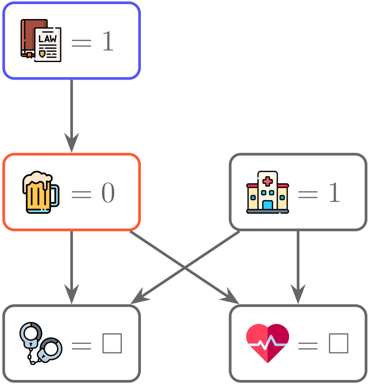
$$X = \ 1\ -\ 0\ , \qquad (F_X)$$

- Strategy for `m`

$$X = 0 \qquad (F_X)$$

- Strategy for `p`

$$X = 1 \qquad (F_X)$$



Example: Exogenous Choice

- Uncontrolled Variables: X, X, X

$$X = 1 \qquad (F_X)$$

$$X = 0 \qquad (F_X)$$

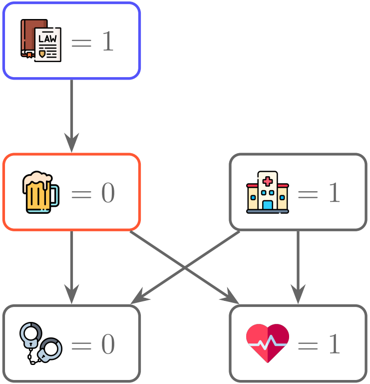
$$X = 1 \qquad (F_X)$$

- Strategy for `m`

$$X = 0 \qquad (F_X)$$

- Strategy for `p`

$$X = 1 \qquad (F_X)$$



Example

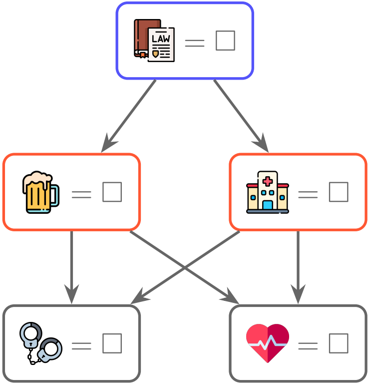
			$u_{\mathtt{m}}$	$u_{\mathtt{p}}$
0	0	-1	2	-1
0	0	0	3	0
0	0	1	4	1
0	1	-1	4	-1
0	1	0	5	0
0	1	1	6	1
1	0	-1	-1	-1
1	0	0	0	0
1	0	1	1	1
1	1	-1	1	-1
1	1	0	2	0
1	1	1	3	1

Example: Endogenous Choice

- Uncontrolled Variables: $X_{\text{hospital}}, X_{\text{handcuffs}}, X_{\text{heart}}$

$$X_{\text{handcuffs}} = \min\{X_{\text{hospital}}, X_{\text{beer}}, X_{\text{law}}\} \qquad (F_{X_{\text{handcuffs}}})$$

$$X_{\text{heart}} = X_{\text{hospital}} - X_{\text{beer}}, \qquad (F_{X_{\text{heart}}})$$



Example: Endogenous Choice

- Uncontrolled Variables: $X_{🏥}, X_{🔗}, X_{❤}$

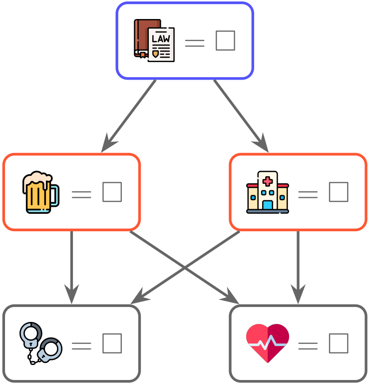
$$X_{🔗} = \min\{X_{🏥}, X_{🍺}, X_{⚖}\} \qquad (F_{X_{🔗}})$$

$$X_{❤} = X_{🏥} - X_{🍺}, \qquad (F_{X_{❤}})$$

- Strategy for `m`

$$X_{🍺} = 1 \qquad (F_{X_{🍺}})$$

$$X_{🏥} = 1 - X_{⚖} \qquad (F_{X_{🏥}})$$



Example: Endogenous Choice

- Uncontrolled Variables: $X_{\text{hospital}}, X_{\text{handcuffs}}, X_{\text{heart}}$

$$X_{\text{handcuffs}} = \min\{X_{\text{hospital}}, X_{\text{beer}}, X_{\text{law}}\} \qquad (F_{X_{\text{handcuffs}}})$$

$$X_{\text{heart}} = X_{\text{hospital}} - X_{\text{beer}}, \qquad (F_{X_{\text{heart}}})$$

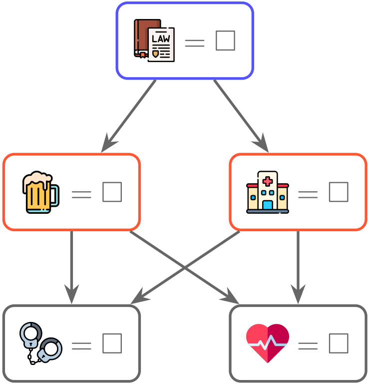
- Strategy for m

$$X_{\text{beer}} = 1 \qquad (F_{X_{\text{beer}}})$$

$$X_{\text{hospital}} = 1 - X_{\text{law}} \qquad (F_{X_{\text{hospital}}})$$

- Strategy for p (mistaken)

$$X_{\text{law}} = 1 \qquad (F_{X_{\text{law}}})$$



Example: Endogenous Choice

- Uncontrolled Variables: $X_{\text{hospital}}, X_{\text{handcuffs}}, X_{\text{heart}}$

$$X_{\text{handcuffs}} = \min\{X_{\text{hospital}}, X_{\text{beer}}, \ 1 \ \} \qquad (F_{X_{\text{handcuffs}}})$$

$$X_{\text{heart}} = X_{\text{hospital}} - X_{\text{beer}}, \qquad (F_{X_{\text{heart}}})$$

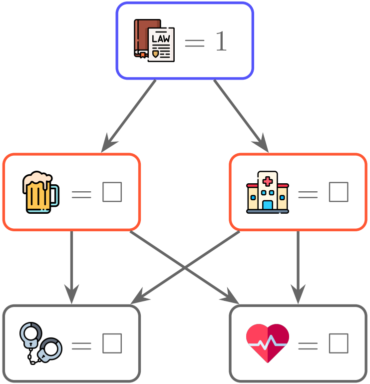
- Strategy for m

$$X_{\text{beer}} = 1 \qquad (F_{X_{\text{beer}}})$$

$$X_{\text{hospital}} = 1 - 1 \qquad (F_{X_{\text{hospital}}})$$

- Strategy for p (mistaken)

$$X_{\text{law}} = 1 \qquad (F_{X_{\text{law}}})$$



Example: Endogenous Choice

◇ Uncontrolled Variables: $X_{🏥}, X_{⛓}, X_{❤}$

$$X_{⛓} = \min\{\ 0\ \ ,\ \ 1\ \ ,\ \ 1\ \ \} \qquad (F_{X_{⛓}})$$

$$X_{❤} = \ \ 0\ \ -\ \ 1\ \ , \qquad (F_{X_{❤}})$$

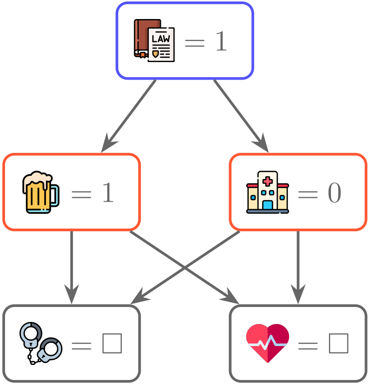
◇ Strategy for m

$$X_{🍺} = 1 \qquad (F_{X_{🍺}})$$

$$X_{🏥} = 0 \qquad (F_{X_{🏥}})$$

◇ Strategy for p (mistaken)

$$X_{📜} = 1 \qquad (F_{X_{📜}})$$



Example: Endogenous Choice

- Uncontrolled Variables: $X_{🏥}, X_{🔗}, X_{❤️}$

$$X_{🔗} = 0 \qquad (F_{X_{🔗}})$$

$$X_{❤️} = -1 \qquad (F_{X_{❤️}})$$

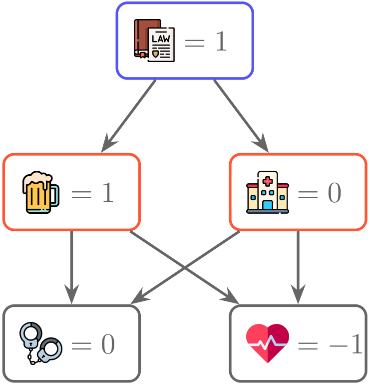
- Strategy for m

$$X_{🍺} = 1 \qquad (F_{X_{🍺}})$$

$$X_{🏥} = 0 \qquad (F_{X_{🏥}})$$

- Strategy for p (mistaken)

$$X_{📜} = 1 \qquad (F_{X_{📜}})$$



Example

			$u_{\mathtt{m}}$	$u_{\mathtt{p}}$
0	0	-1	2	-1
0	0	0	3	0
0	0	1	4	1
0	1	-1	4	-1
0	1	0	5	0
0	1	1	6	1
1	0	-1	-1	-1
1	0	0	0	0
1	0	1	1	1
1	1	-1	1	-1
1	1	0	2	0
1	1	1	3	1

In the paper:

- We allow for probabilistic structural equations
- Consider how player’s update beliefs when surprised
- What kind of misperception is sustainable in the long run?

Thanks