A REPRESENTATION THEOREM FOR CAUSAL DECISION MAKING

Evan Piermont
Royal Holloway, University of London

joint with

Joseph Y. Halpern
Cornell University

Aarhus Workshop --- May 2024

This paper

We represent causality via *structural equations*, and consider an agent's preference over *interventions*:

- Representation Theorem
  - How does an agent's subjective causal model influence her decision making
- Identification Theorem
  - When can this model be recovered from observation

Causation and Counterfactuals

- Modern theories define causation through counterfactuals.
    - Requires evaluating worlds that do not exist
- We take a structural approach a la Pearl [2000]:
    - Equations directly encode causal mechanisms
    - Provide a succinct way of contemplating counterfactuals

Variables

- $\mathcal{U}$ and $\mathcal{V}$ denote exogenous and endogenous variables, resp.
- $\mathcal{R}(Z) \subset \mathbb{R}$ is the range of $Z \in \mathcal{U} \cup \mathcal{V}$
- A *context* is a vector $\vec{u}$ of values for all the exogenous variables $\mathcal{U}$.
  - Let $\texttt{ctx} = \prod_{U \in \mathcal{U}} \mathcal{R}(U)$ collect all contexts
- A *resolution* is a vector $\vec{r}$ of values for all variables $\mathcal{U} \cup \mathcal{V}$.
  - Let $\texttt{res} = \prod_{Y \in \mathcal{U} \cup \mathcal{V}} \mathcal{R}(Y)$ collect all resolutions

The decision maker cares about the resolution and is uncertain about the context:

- Utility will be defined over `res`
- Beliefs will be defined over `ctx`

Example

The US Federal Reserve is contemplating the economy.

The relevant variables are: the growth rate (gw), the prior interest rate (pr), the current interest rate (rate), inflation (inf), employment rate (emp):

$$\mathcal{U} = \begin{cases} U_{gw} \\ U_{pr} \end{cases} \qquad \mathcal{V} = \begin{cases} Y_{rt} \\ X_{emp} \\ X_{inf} \end{cases}$$

Example

- Utility is determined by the inflation rate and employment level:
  - $\mathbf{u}(\vec{r}) = 2X_{emp} - X_{inf}$.
- Does not know the growth rate:
  - believes $U_{gw} = 1$ with prob α and $U_{gw} = 0$ with prob $(1-\alpha)$.
- Contemplate interventions that set the interest rate:
  - This will casually effect the resolution
  - But exactly how might depend on the context

Causal Models

Given $\mathcal{U}$ and $\mathcal{V}$ with ranges $\mathcal{R}$, a **causal model M** consists of:

- $\mathcal{F} = \{F_X\}_{X \in \mathcal{V}}$, a set of **structural equations**, where

$$F_X : \prod_{Y \in \mathcal{U} \cup (\mathcal{V} - \{X\})} \mathcal{R}(Y) \to \mathcal{R}(X).$$

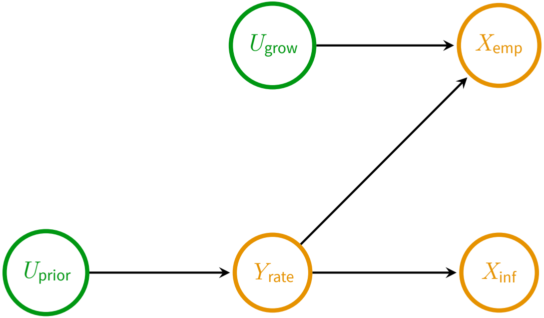
- Call **M** *recursive* if there exists a partial order on $\mathcal{V}$:
  - F_X is independent of the variables succeeding X

Example

The causal equations are

$$
\begin{aligned}
Y_{rt} &= U_{pr} && (F_{Y_{rt}}) \\
X_{inf} &= 1 - Y_{rt} && (F_{X_{inf}}) \\
X_{emp} &= 1 - (Y_{rt} \times (1 - U_{gw})) && (F_{X_{inf}})
\end{aligned}
$$

Example



Example

Given a recursive **M**, each context $\vec{u}$ induces a unique resolution $\vec{r}$:

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= U_{pr} \\
X_{inf} &= 1 - Y_{rt} \\
X_{emp} &= 1 - (Y_{rt} \times (1 - U_{gw}))
\end{aligned}
$$

Example

Given a recursive **M**, each context $\vec{u}$ induces a unique resolution $\vec{r}$:

$$\begin{aligned} U_{gw} &= 0 \\ U_{pr} &= 0 \\ Y_{rt} &= 0 \\ X_{inf} &= 1 - Y_{rt} \\ X_{emp} &= 1 - (Y_{rt} \times (1 - 0)) \end{aligned}$$

Example

Given a recursive **M**, each context $\vec{u}$ induces a unique resolution $\vec{r}$:

$$\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= 0 \\
X_{inf} &= 1 - 0 \\
X_{emp} &= 1 - (0 \times (1 - 0))
\end{aligned}$$

Example

Given a recursive **M**, each context $\vec{u}$ induces a unique resolution $\vec{r}$:

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= 0 \\
X_{inf} &= 1 \\
X_{emp} &= 0
\end{aligned}
$$

- When is decision making consistent causal reasoning via some model **M**?
- What kind of data is needed to answer this?
- Preferences over **interventions**

Interventions

An **intervention**

$$\mathbf{do}[\,Y_1 \leftarrow y_1, \ldots, Y_n \leftarrow y_n]$$

is a mediation that sets the values of $Y_1 \ldots Y_n \in \mathcal{V}$:

- $y_i \in \mathcal{R}(Y_i)$
- abbreviated as $\mathbf{do}[\,\vec{Y} \leftarrow \vec{y}]$
- interventions only on endogenous variables.

A **conditional intervention** is of the form:

$$\textbf{if } \phi \textbf{ then } A \textbf{ else } B$$

- ϕ is a true/false valued question about the variable values
  - such as “the value of X is positive”, etc
- A and B are conditional interventions
- These is constructed recursively starting with interventions
- **if** ϕ **then** A shorthand for when $B = \emptyset$

For a resolution $\vec{r} \in \texttt{res}$ let

$$\textbf{if} \;\; \vec{r} \;\; \textbf{then} \;\; A \;\; \textbf{else} \;\; B$$

denote the conditional intervention on $\vec{r}$ being true.

- i.e., do A if all variables coincide with $\vec{r}$, else do B

Preference

Observable: preference relation $\succsim$ over conditional interventions:

- Interventions allow the DM to change the resolution
- Conditioning allows contracting away uncertainty about context

A **causally sophisticated** decision maker would understand the effect of conditional interventions via a causal model

Interventions and Causal models

Given the model **M**, the intervention

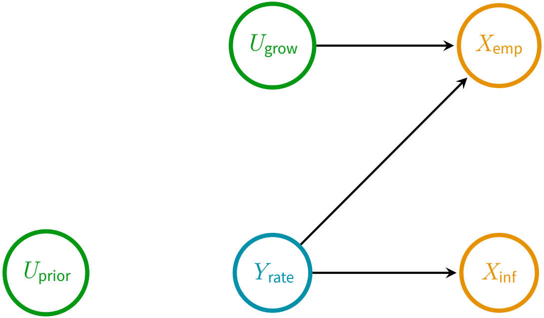
$$\mathbf{do}[\,Y_1 \leftarrow y_1, \ldots, Y_n \leftarrow y_n]$$

induces a *counterfactual model*, $\mathcal{F}_{\mathbf{do}[\,\vec{Y} \leftarrow \vec{y}]}$ where

F_{Y_i} is replaced by the constant function $F'_{Y_i} = y_i$

Example

The intervention $\mathbf{do}[Y_{rt} \leftarrow 1]$ sets the current rate to 1:



Example

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= 1 \\
X_{inf} &= 1 - Y_{rt} \\
X_{emp} &= 1 - (Y_{rt} \times (1 - U_{gw}))
\end{aligned}
$$

Example

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= 1 \\
X_{inf} &= 1 - Y_{rt} \\
X_{emp} &= 1 - (Y_{rt} \times (1 - 0))
\end{aligned}
$$

Example

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= 1 \\
X_{inf} &= 1 - 1 \\
X_{emp} &= 1 - (1 \times (1 - 0))
\end{aligned}
$$

Example

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= 1 \\
X_{inf} &= 0 \\
X_{emp} &= 0
\end{aligned}
$$

Given a (recursive) model **M** and conditional intervention A, let

$$\beta_A^{\mathbf{M}} : \mathtt{ctx} \to \mathtt{res}$$

transform contexts into resolutions in the obvious way:

- **M** plus context determines ex-ante resolution
- This resolution determines the 'clause' of A in force, hence an intervention
- This intervention determines a (recursive) counterfactual model
- Along with context, this determines the ex-post resolution

Representation

A causally sophisticated agent's preferences are parameterized by

- **M** — a recursive model capturing causal relationships
- $\mathbf{u} : \mathtt{res} \to \mathbb{R}$ — value of a resolution of all uncertainty
- $\mathbf{p} \in \Delta(\mathtt{ctx})$ — belief capturing uncertainty about the values of exogenous (hence endogenous) variables

Representation

Subjective Causal Utility

$(\mathbf{M}, \mathbf{p}, \mathbf{u})$ is a **subjective causal utility representation** of $\succsim$:

$$A \succsim B$$

if and only if

$$\sum_{\vec{u} \in \mathtt{ctx}} \mathbf{u}(\beta_A^{\mathbf{M}}(\vec{u}))\mathbf{p}(\vec{u}) \geq \sum_{\vec{u} \in \mathtt{ctx}} \mathbf{u}(\beta_B^{\mathbf{M}}(\vec{u}))\mathbf{p}(\vec{u}).$$

Example

The utility of the Federal Reserve is determined by the inflation rate and employment level, and is given by

$$\mathbf{u}(\vec{r}) = 2X_{emp} - X_{inf}.$$

Example

$$\mathbf{do}[\,Y_{rt} \leftarrow 0]$$

$$\begin{tikzcd}
& \begin{cases} Y_{rt} & = 0 \\ X_{inf} & = 1 \\ X_{emp} & = 1 \end{cases} \\
\beta \arrow[ur, "U_{gw} = 1"] \arrow[dr, "U_{gw} = 0"'] & \\
& \begin{cases} Y_{rt} & = 0 \\ X_{inf} & = 1 \\ X_{emp} & = 1 \end{cases}
\end{tikzcd}$$

$$\mathbf{do}[\,Y_{rt} \leftarrow 1]$$

$$\begin{tikzcd}
& \begin{cases} Y_{rt} & = 1 \\ X_{inf} & = 0 \\ X_{emp} & = 1 \end{cases} \\
\beta \arrow[ur, "U_{gw} = 1"] \arrow[dr, "U_{gw} = 0"'] & \\
& \begin{cases} Y_{rt} & = 1 \\ X_{inf} & = 0 \\ X_{emp} & = 0 \end{cases}
\end{tikzcd}$$

Example

$$\mathbf{do}[\,Y_{rt} \leftarrow 0]$$

$$\begin{tikzcd}
 & 2 \times 1 - 1 = 1 \\
\mathbf{u} \circ \beta \arrow[ur, "U_{gw} = 1"] \arrow[dr, "U_{gw} = 0"'] & \\
 & 2 \times 1 - 1 = 1
\end{tikzcd}$$

Utility is 1

$$\mathbf{do}[\,Y_{rt} \leftarrow 1]$$

$$\begin{tikzcd}
 & 2 \times 1 - 0 = 2 \\
\mathbf{u} \circ \beta \arrow[ur, "U_{gw} = 1"] \arrow[dr, "U_{gw} = 0"'] & \\
 & 2 \times 0 - 0 = 0
\end{tikzcd}$$

Utility is 2α

Example

- Preference between setting interest rate at 1 or 0 depends on belief about U_{gw}.

- The conditional intervention

$$\textbf{if}\,(U_{gw} = 1)\ \textbf{then}\ \textbf{do}[Y_{rt} \leftarrow 1]\ \textbf{else}\ \textbf{do}[Y_{rt} \leftarrow 0]$$

  dominates

Axioms

We obtain the additive structure via a `cancellation` axiom

- Adapted from Blume, Easley, Halpern (2021)
- In the spirit of Krantz, Luce, Suppes & Tversky (1971)

Call $\vec{r}$ **null** if

$$(\textbf{if } \vec{r} \textbf{ then } A) \sim (\textbf{if } \vec{r} \textbf{ then } B) \qquad \text{for all } A \text{ and } B$$

- Conditioning on $\vec{r}$ trivializes preference
- The DM does not believe $\vec{r}$ is possible

Ax 1: Model Uniqueness

For each $\vec{u} \in \texttt{ctx}$, there is at most one $\vec{r} \in \texttt{res}$ such that $\vec{r}|_{\mathcal{U}} = \vec{u}$ and $\vec{r}$ is non-null.

- For each context, there is at most one consistent resolution considered possible
- i.e., given the context, there is no uncertainty about the resolution
- Implies the casual model is certain

For each $\vec{r} \in \texttt{res}$, write

$$\mathbf{do}[\vec{Y} \leftarrow \vec{y}] \rightsquigarrow_{\vec{r}} (X = x)$$

as shorthand for the indifference relation

$$\textbf{if } \vec{r} \textbf{ then } \mathbf{do}[\vec{Y} \leftarrow \vec{y}, X \leftarrow x] \sim \textbf{if } \vec{r} \textbf{ then } \mathbf{do}[\vec{Y} \leftarrow \vec{y}].$$

- If setting $\vec{Y}$ to $\vec{y}$ yields $X = x$, then the agent is indifferent from making such a further intervention on X.
- However, definition allows for indifference between distinct values of X

Example

No Intervention

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= U_{pr} \\
X_{inf} &= 1 - Y_{rt} \\
X_{emp} &= 1 - (Y_{rt} \times (1 - U_{gw}))
\end{aligned}
$$

$\mathbf{do}[Y_{rt} \leftarrow 1]$

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= 1 \\
X_{inf} &= 1 - Y_{rt} \\
X_{emp} &= 1 - (Y_{rt} \times (1 - U_{gw}))
\end{aligned}
$$

$\mathbf{do}[Y_{rt} \leftarrow 1, X_{emp} \leftarrow 0]$

$$
\begin{aligned}
U_{gw} &= 0 \\
U_{pr} &= 0 \\
Y_{rt} &= 1 \\
X_{inf} &= 1 - Y_{rt} \\
X_{emp} &= 0
\end{aligned}
$$

Example

No Intervention	$\mathbf{do}[Y_{rt} \leftarrow 1]$	$\mathbf{do}[Y_{rt} \leftarrow 1, X_{emp} \leftarrow 0]$
$U_{gw} = 0$	$U_{gw} = 0$	$U_{gw} = 0$
$U_{pr} = 0$	$U_{pr} = 0$	$U_{pr} = 0$
$Y_{rt} = 0$	$Y_{rt} = 1$	$Y_{rt} = 1$
$X_{inf} = 1 - Y_{rt}$	$X_{inf} = 1 - Y_{rt}$	$X_{inf} = 1 - Y_{rt}$
$X_{emp} = 1 - (Y_{rt} \times (1 - 0))$	$X_{emp} = 1 - (Y_{rt} \times (1 - 0))$	$X_{emp} = 0$

Example

No Intervention

$$U_{gw} = 0$$
$$U_{pr} = 0$$
$$Y_{rt} = 0$$
$$X_{inf} = 1 - 0$$
$$X_{emp} = 1 - (0 \times (1 - 0))$$

$\mathbf{do}[Y_{rt} \leftarrow 1]$

$$U_{gw} = 0$$
$$U_{pr} = 0$$
$$Y_{rt} = 1$$
$$X_{inf} = 1 - 1$$
$$X_{emp} = 1 - (1 \times (1 - 0))$$

$\mathbf{do}[Y_{rt} \leftarrow 1, X_{emp} \leftarrow 0]$

$$U_{gw} = 0$$
$$U_{pr} = 0$$
$$Y_{rt} = 1$$
$$X_{inf} = 1 - 1$$
$$X_{emp} = 0$$

Example

No Intervention	$\mathbf{do}[Y_{rt} \leftarrow 1]$	$\mathbf{do}[Y_{rt} \leftarrow 1, X_{emp} \leftarrow 0]$
$U_{gw} = 0$	$U_{gw} = 0$	$U_{gw} = 0$
$U_{pr} = 0$	$U_{pr} = 0$	$U_{pr} = 0$
$Y_{rt} = 0$	$Y_{rt} = 1$	$Y_{rt} = 1$
$X_{inf} = 1$	$X_{inf} = 0$	$X_{inf} = 0$
$X_{emp} = 1$	$X_{emp} = 0$	$X_{emp} = 0$

Ax 2: Definiteness

Fix non-null $\vec{r} \in \texttt{res}$, endogenous variables, $\vec{Y}$, and values $\vec{y} \in \mathcal{R}(\vec{Y})$. Then for variable X, there exists some $x \in \mathcal{R}(X)$ such that

$$\mathbf{do}[\vec{Y} \leftarrow \vec{y}] \rightsquigarrow_{\vec{r}} (X = x)$$

- There is some value of X which is consistent with any intervention
- May not be unique (i.e., indifference between resolutions)
- `Ax2`*: if the value x is unique

Ax 3: Centeredness

For $\vec{r} \in \texttt{res}$, vector of endogenous variables $\vec{Y}$, and endogenous variable $X \notin \vec{Y}$, we have

$$\mathbf{do}[\vec{Y} \leftarrow \vec{r}|_{\vec{Y}}] \rightsquigarrow_{\vec{r}} (X = \vec{r}|_X)$$

- Trivial interventions (setting variables to their current value) has no consequence

For $X, Y \in \mathcal{V}$, say that X *is unaffected by* Y if

$$\mathbf{do}[\vec{Z} \leftarrow \vec{z}] \rightsquigarrow_{\vec{r}} (X = x) \qquad \text{iff} \qquad \mathbf{do}[\vec{Z} \leftarrow \vec{z}, Y \leftarrow y] \rightsquigarrow_{\vec{r}} (X = x)$$

for all $\vec{r} \in \texttt{res}$, $\vec{Z}$ and values for the variables.

- X is unaffected by Y if there is no intervention on Y that changes the decision maker's perception of X
- If this relation does not hold, then X is *affected* by Y, written $Y \rightsquigarrow X$.

Ax 4: Recursivity

$\rightsquigarrow$ is acyclic

- There are no cycles of variable dependence

Theorem

$\succsim$ satisfies `Ax1-4` and `cancellation` if and only if there exists a subjective causal utility representation, $(\mathbf{M}, \mathbf{p}, \mathbf{u})$.

Moreover, if $\texttt{Ax2}^*$ holds, then $\mathbf{M}$ is unique.

Each axiom helps discipline how counterfactuals are constructed:

`Definiteness:` There exists some counterfactual world

`Model Uniqueness:` It is unique

`Centeredness:` It is minimally different than the current world

`Recursivity:` Closeness is consistent across contexts

These properties suffice to prove the existence of a structural model.

Thank You!