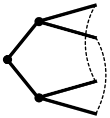
Causality and Causal Misperception in Dynamic Games

Sungmin Park

The Ohio State University

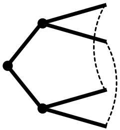
May 18, 2024

Early work welcoming comments



What I do

Question	What is a useful way to model people's misperceptions about causal relationships?
Answer	Let agents have observation-consistent expectations (OCE) or Maximum Entropy (MaxEnt) OCE



Motivation

Difficulty of correct causal inference

Inferring causality from observed data is difficult

- Difficulty is due to unobserved variables or simultaneity, e.g.:
  - What is the effect of education on earnings?
  - What is the effect of police on crime?
- The challenge persists even as the sample size grows large
- Much work by econometricians, applied microeconomists, statisticians, and computer scientists is to address this hurdle

Given this difficulty, why should we expect agents in our models to have correct beliefs about causality?

Books on causal inference

Main results

OCE and **MaxEnt OCE**	An observation-consistent expectations (OCE) is maximum entropy (MaxEnt) OCE if and only if it exhibits correlation neglect
MaxEnt OCE Equilibrium	Every finite extensive-form game with perfect recall and observational constraint has a MaxEnt OCE equilibrium.
Causality (Not today)	A causal relation satisfies the axioms of causation if and only if it has a probabilistic "event structure representation."

Literature

Decision making under causal misperceptions

- **Theory:** DM's perception is distorted by a subjective DAG* that is exogenous (Spiegler, 2016, 2022, 2023) or chosen by DM (Eliaz and Spiegler, 2020; Eliaz et al., 2022)
  - * A directed acyclic graph (DAG) specifies a set of conditional independence assumptions between random variables.
- **Experiment:** When subjects are given the same data but are presented with different causal narratives, they make different choices (Kendall and Charles, 2022)

Self-confirming equilibrium (SCE) and conjectural equilibrium (CE)

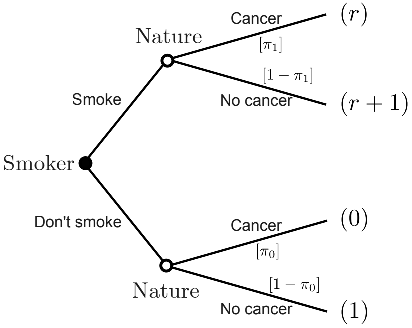
- **SCE**: Each player has a correct belief about others' strategies on the equilibrium path of play (Fudenberg and Levine, 1993; Fudenberg and Kreps, 1995)
- **CE**: Each player has a belief about others' strategies consistent with observation (Battigalli and Guaitoli, 1988; Battigalli, 1997; Azrieli, 2009)

Simplest example

And a silly one; smokers please don't take this seriously

A smoker's decision problem

- Smoker chooses to smoke ($s = 1$) or not ($s = 0$)
- Nature gives cancer ($y = 1$) with probability π_s and no cancer ($y = 0$) with probability $1 - \pi_s$
  - Smoking **causes** cancer: $\pi_1 > \pi_0$
- Smoker gets r for smoking and 1 for staying healthy
- A **strategy** is the probability $\sigma \in [0, 1]$ of smoking.
- A smoker's **belief** is $\beta = (\beta_0, \beta_1)$ where β_s is the subjective probability of getting cancer with s.



OCE

Definition (OCE)

Given strategy $\sigma \in [0,1]$, an **observation-consistent expectations (OCE)** is a belief $\beta \in [0,1]^2$ such that

$$(1-\sigma)\beta_0 + \sigma\beta_1 = (1-\sigma)\pi_0 + \sigma\pi_1.$$

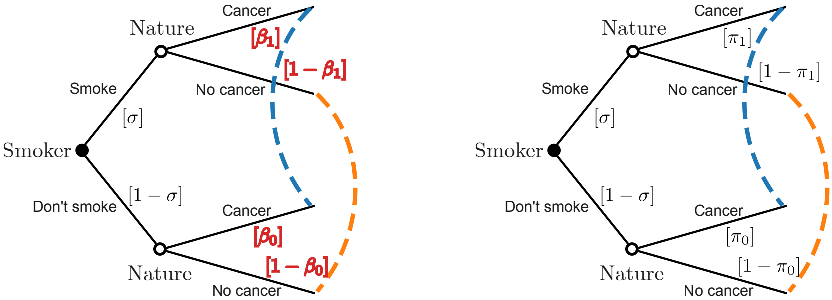
Interpretation

- Smoker sees others choosing σ and getting cancer with frequencey $\sigma\pi_1 + (1-\sigma)\pi_0$, but does not know π_0 or π_1
- What the smoker thinks Nature does (β) and what Nature really does (π) are observationally equivalent

Illustration of OCE

Observational equivalence

What I **think** Nature does | What Nature **really** does



Remark: There are infinitely many OCEs for any strategy $\sigma \in [0, 1]$.

MaxEnt OCE

Let $\mathbf{p}(\sigma, \beta)$ be the vector of probabilities over the 4 outcomes (s, y):

$$\mathbf{p}(\sigma, \beta) = [\sigma\beta_1 \quad \sigma(1-\beta_1) \quad (1-\sigma)\beta_0 \quad (1-\sigma)(1-\beta_0)]^T.$$

Definition (MaxEnt OCE)

Given strategy $\sigma \in (0, 1)$, an OCE $\beta^* \in [0, 1]^2$ is a **maximum entropy (MaxEnt) OCE** if it satisfies

$$\beta^* \in \underset{\beta \in [0,1]^2}{\operatorname{argmax}} \; G(\mathbf{p}(\sigma, \beta))$$

subject to $\qquad (1-\sigma)\beta_0 + \sigma\beta_1 = (1-\sigma)\pi_0 + \sigma\pi_1, \qquad$ (β is an OCE)

where $G(\cdot)$ is the Shannon entropy function.

Interpretation

- MaxEnt OCE is the belief with the least information among all OCEs

Result

MaxEnt OCE $\Rightarrow$ correlation neglect

Proposition

For every $\sigma \in (0,1)$, the MaxEnt OCE β^* satisfies

$$\beta_0^* = \beta_1^* = (1-\sigma)\pi_0 + \sigma\pi_1.$$

Meaning. The smoker doesn't think smoking causes cancer

Intuition. The smoker observes no evidence of dependence between smoking and cancer, so he believes in none.

General result (Theorem 1). MaxEnt OCE $\Leftrightarrow$ correlation neglect, whenever agents observe only the marginal prob. distribution between two variables

Definition of equilibrium

Motivation

- OCE and MaxEnt OCE take the strategy σ as given
- Is there an equilibrium where this σ is subjectively optimal?

Definition

A strategy-belief pair (σ, β) is an **OCE equilibrium** if

1. Given the belief β, the strategy σ is a best response, and
2. Given the strategy σ, the belief β is an OCE.

An OCE equilibrium (σ, β) is a **MaxEnt OCE equilibrium** if some $\{(\sigma^j, \beta^j)\}_{j=1}^{\infty} \to (\sigma, \beta)$ where each σ^j is the MaxEnt OCE given β.

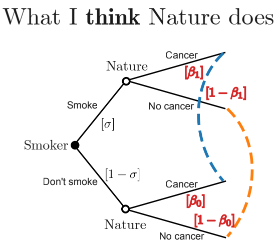
Result: OCE equilibria

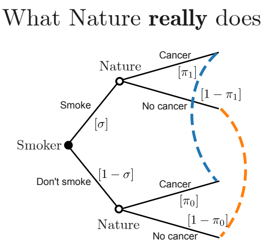
Every strategy is rationalizable with some OCE

Proposition

A strategy-belief pair (σ, β) *is an OCE equilibrium if and only if*

1. $\sigma = 0$, $\beta_0 = \pi_0$, *and* $\beta_1 \geq \pi_0 + r$,
2. $\sigma = 1$, $\beta_1 = \pi_1$, *and* $\beta_0 \geq \pi_1 - r$, *or*
3. $\sigma \in (0, 1)$, $\beta_0 = (1 - \sigma)\pi_0 + \sigma(\pi_1 - r)$, *and* $\beta_1 = (1 - \sigma)(\pi_0 + r) + \sigma\pi_1$.





Result: MaxEnt equilibrium

A sharper prediction

Proposition

A strategy-belief pair (σ, β) is a MaxEnt OCE equilibrium if and only if

$$\sigma = 1 \quad \textit{and} \quad (\beta_0, \beta_1) = (\pi_1, \pi_1).$$

Meaning

- Continue smoking while thinking that smoking doesn't cause cancer

Intuition

- MaxEnt OCE implies correlation neglect, so no other strategy is a best response.

Generalizing the observational constraint

Motivation

- Correlation neglect sounds too naive. Can we make agents more sophisticated? Yes! Give them better observation

Definition

Given an **observation constraint matrix** C and strategy σ, an **OCE** is a belief β such that

$$C\mathbf{p}(\sigma, \beta) = C\mathbf{p}(\sigma, \pi).$$

Examples of C:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

Generalizing the approach to extensive-form games

Finite extensive-form game with perfect recall + observational constraint

- N: set of players,
- H: set of histories,
- ι: mapping of non-terminal histories to players,
- π: probability distribution of Nature's moves,
- $\mathcal{I}$: collection of information sets,
- u: payoff function, and
- C: **observational constraint matrix**

OCE and MaxEnt OCE

- σ_i: Player i's (behavioral) strategy
- β_i: Player i's ex-ante beliefs about others' moves (including Nature's)

Definition

Let a strategy σ_i be given. A belief β_i is an **OCE for player i** if

$$C\mathbf{p}(\sigma_i, \beta_i) = C\mathbf{p}(\sigma_i, (\pi, \sigma_{-i})).$$

Let a totally mixed strategy σ_i be given. An OCE for player i β_i^* is a **MaxEnt OCE for player i** if

$$\beta^* \in \operatorname*{argmax}_{\beta_i \text{ is an OCE}_i} G(\mathbf{p}(\sigma_i, \beta_i)),$$

where G is the Shannon entropy function.

Equilibrium

- μ_i: Player i's ex-post beliefs about histories within information sets

Definition

A triple (σ, β, μ) is an **OCE equilibrium** if for every player i,

1. **(Sequential rationality)** the strategy σ_i is sequentially rational given (β_i, μ_i):

   $\sigma_i \in \operatorname{argmax} u_i(\sigma_i, \beta_i, \mu_i | I)$ at every info set $I \in \mathcal{I}_i$

2. **(Observational consistency)** the ex-ante belief β_i is an OCE given σ_i, and

3. **(Bayes-consistency)** the ex-post belief μ_i satisfies Bayes rule given (σ_i, β_i).

An OCE equilibrium (σ, β, μ) is a **MaxEnt equilibrium** if some $\{(\sigma^k, \beta^k)_{k=1}^{\infty}\} \to (\sigma, \beta)$ where each σ_i^k is MaxEnt OCE given β_i^k.

Result: existence

Theorem

Every finite extensive-form game with perfect recall and observational constraint has a MaxEnt OCE equilibrium.

Meaning

- There always exists an equilibrium where everyone is best-responding to their misperceptions, even as those beliefs are the least crazy ones based on observation.

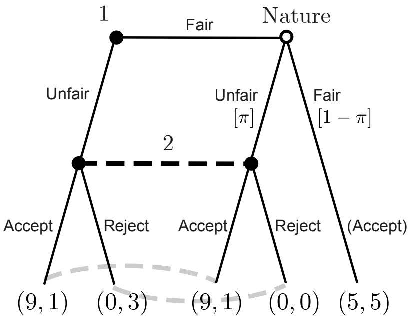
Proof idea

- With ϵ-constrained strategies, mappings from a strategy profile σ to a MaxEnt OCE β and ex post belief μ are well-behaved.

Illustration: Ultimatum-like game with causal misperception

Manager-Worker game

- Manager (Player 1) decides a fair or unfair bonus to Worker (Player 2)
- Even if manager chooses a fair bonus, Nature might change it to unfair or keep it fair
- If worker receives fair bonus, he accepts. If not, he either accepts or rejects.
  - He gets a thrill for rejecting an unfair Manager
- Worker doesn't know how likely manger treats him unfairly ex ante or ex post



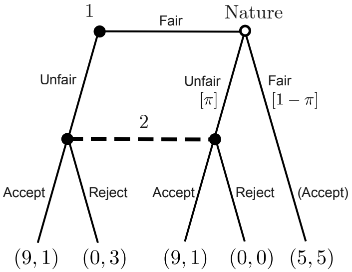
$$C = \begin{bmatrix} 1 & \cdot & 1 & \cdot & \cdot \\ \cdot & 1 & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

Standard prediction

Proposition

In the unique Perfect Bayesian Equilibrium,

- *Manager offers an unfair bonus with probability* $\frac{\pi}{2+\pi}$
- *When Worker gets offered an unfair bonus, he correctly believes it's 67% caused by Nature, and accepts 5 out of 9 times*



Intuition

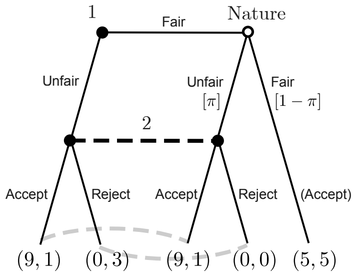
- There is no ex-ante uncertainty about other players' strategies, so there is no causal misperception

My prediction

Proposition

In the unique MaxEnt OCE equilibrium,

- *Manager correctly believes worker will reject if offered unfair bonus. She offers a fair bonus*
- *Worker incorrectly believes that Manager mixes. If offered unfair bonus, he incorrectly thinks that it's caused by either Manager or Nature each with 50% chance, and rejects.*



$$C = \begin{bmatrix} 1 & \cdot & 1 & \cdot & \cdot \\ \cdot & 1 & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

Intuition

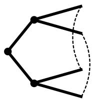
- Worker has no clue about Manager's and Nature's strategies beyond what he observes with C

Takeaways

Use my solution concept if you want to . . .

- allow causal misperception in a dynamic model
- let misperception arise endogenously from observational constraints, and
- want narrow predictions

Thank you!



Appendix

References I

Azrieli, Yaron (2009) "On pure conjectural equilibrium with non-manipulable information," *International Journal of Game Theory*, 38, 209–219.

Battigalli, Pierpaolo (1997) "On rationalizability in extensive games," *Journal of Economic Theory*, 74 (1), 40–61.

Battigalli, Pierpaolo and Danilo Guaitoli (1988) *Conjectural equilibria and rationalizability in a macroeconomic game with incomplete information*: Università commerciale L. Bocconi.

Eliaz, Kfir, Simone Galperti, and Ran Spiegler (2022) "False narratives and political mobilization," *arXiv preprint arXiv:2206.12621*.

Eliaz, Kfir and Ran Spiegler (2020) "A model of competing narratives," *American Economic Review*, 110 (12), 3786–3816.

Fudenberg, Drew and David M Kreps (1995) "Learning in extensive-form games I. Self-confirming equilibria," *Games and Economic Behavior*, 8 (1), 20–55.

References II

Fudenberg, Drew and David K Levine (1993) "Self-confirming equilibrium," *Econometrica: Journal of the Econometric Society*, 523–545.

Kendall, Chad W and Constantin Charles (2022) "Causal narratives,"Technical report.

Spiegler, Ran (2016) "Bayesian networks and boundedly rational expectations," *The Quarterly Journal of Economics*, 131 (3), 1243–1290.

——— (2022) "On the behavioral consequences of reverse causality," *European Economic Review*, 149, 104258.

——— (2023) "Behavioral Causal Inference," *arXiv preprint arXiv:2305.18916*.