Characterizing Optimal Mixed Policies:
Where to Intervene and What to Observe

Sanghack Lee
Elias Bareinboim
Causal AI Laboratory
Columbia University

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

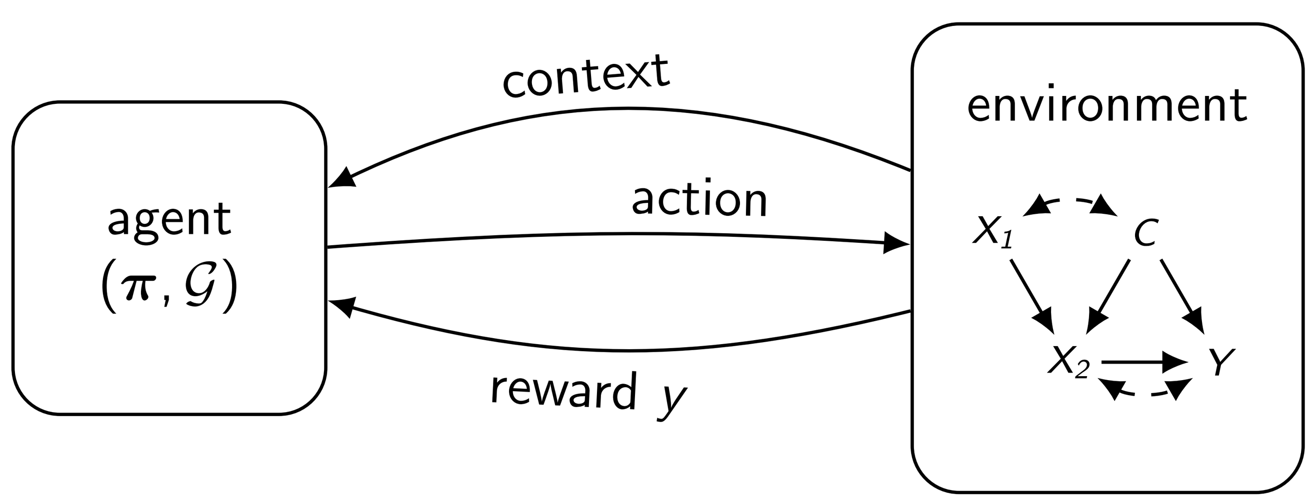
NEURAL INFORMATION PROCESSING SYSTEMS

summary

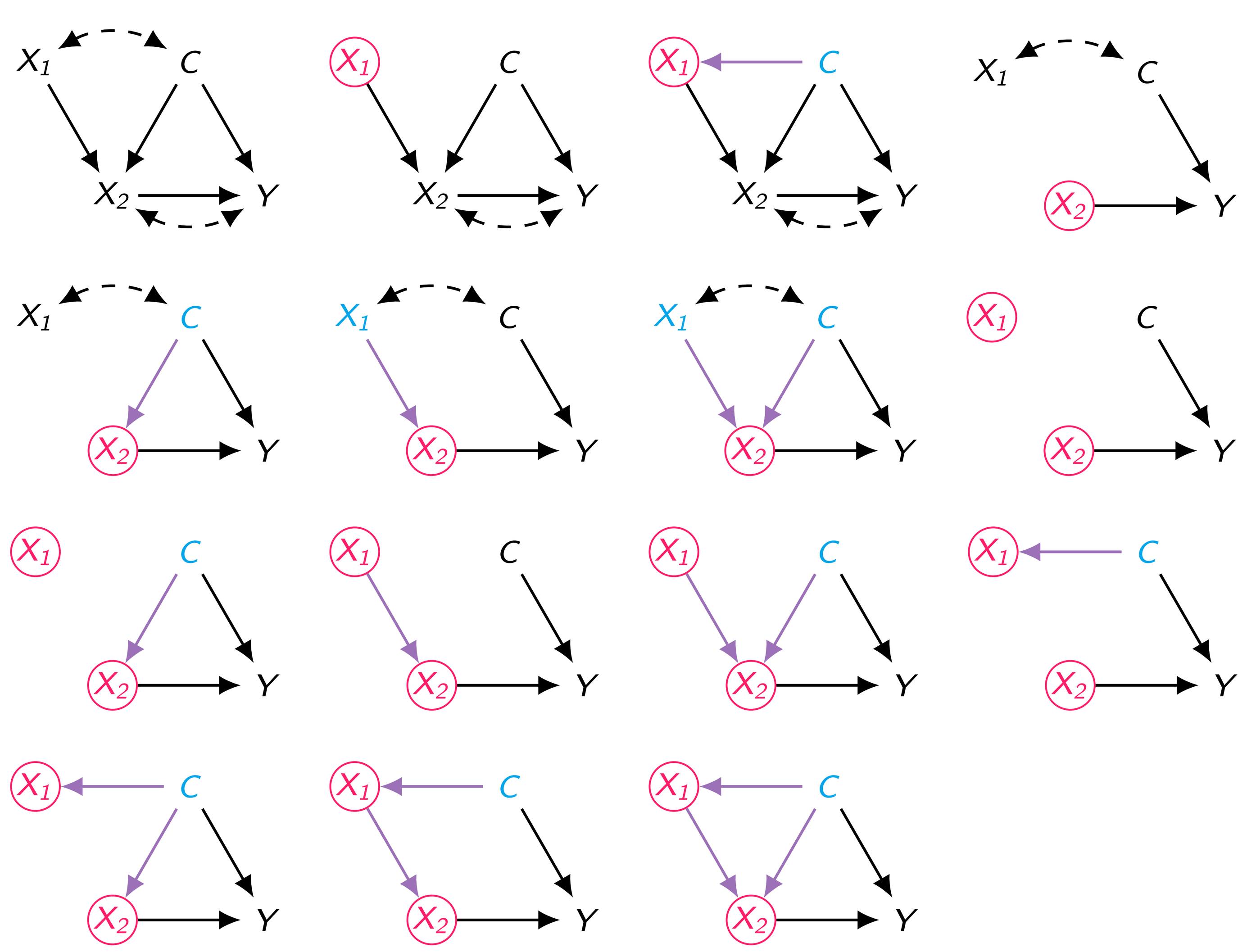
- **Causal understanding** of a system → emergence of a **mixed policy** where an agent can intervene on a **subset** of action variables based on a **subset** of contexts.
- **Challenges**: an agent should figure out which **scopes of mixed policy** (which variables to intervene and observe) to explore to converge **efficiently** and **effectively**.
- **Characterizations**: Given a causal graph of an environment but not its underlying mechanisms, a **scope** is characterized by
  - (**non-redundancy under optimality**): whether contexts or interventions in the scope are **necessary** to obtain an optimal reward.
  - (**possible-optimality**) whether optimizing agent's policy along with the scope converges to **optimal**.
- **Conclusions**: Given a causal graph, an intelligent agent can avoid examining unnecessarily inefficient and ineffective policies so as to converge to optimal faster

an illustrative example

Consider the interaction between an agent and an environment

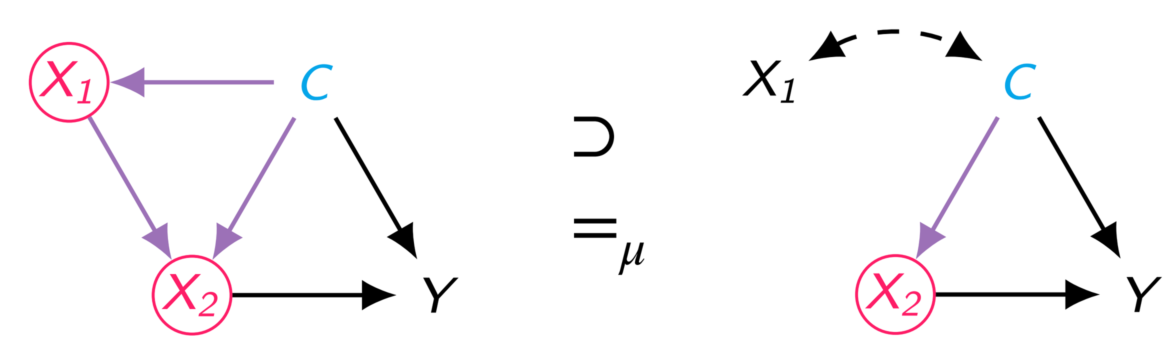


Given **intervenable variables** $\{X_1, X_2\}$ and **observable variables** $\{C, X_1\}$ (~ can be used as contexts), there are **15 different combinations** (**mixed policy scopes, MPS**) for the agent to act upon the system ranging from simple observation to intervening on every variable:



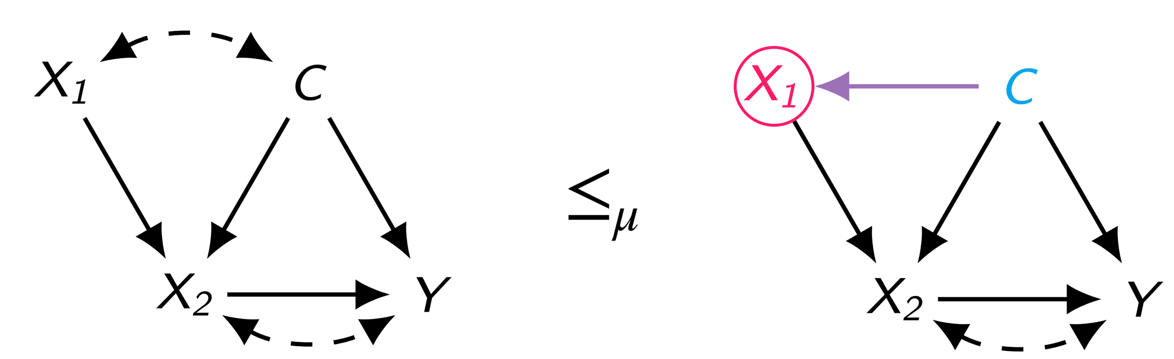
* Intervened variables are circled.
* Variables pointing to an intervened variable are context for the action variable (directed edges onto the intervened variables manifest policy-induced dependency.)

- **non-redundancy**: Given an MPS, no smaller MPS can achieve the same expected rewards in every environment.



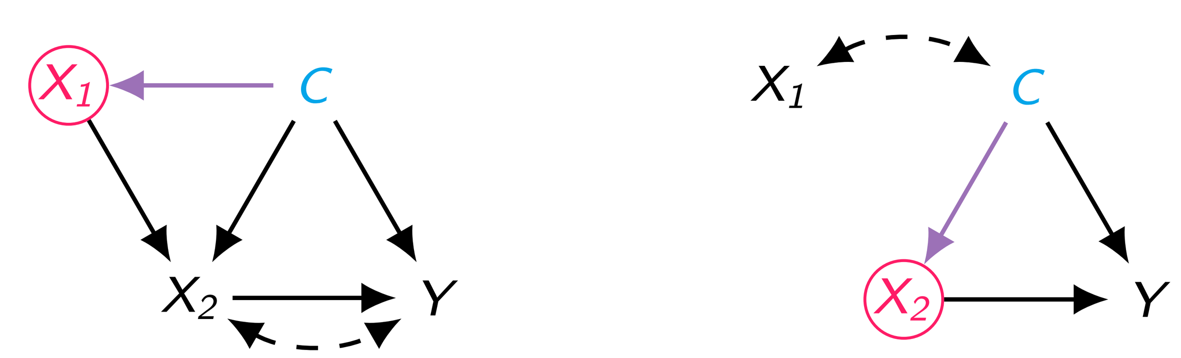
→ the left MPS is *redundant* because the right MPS is smaller ($\supset$, in a sense that having less actions and/or contexts) while being able to attain the same expected reward (denoted by $=_\mu$).

- **optimality**: Given an MPS, there exists no "better than or equal" MPS w.r.t. an optimal expected reward.



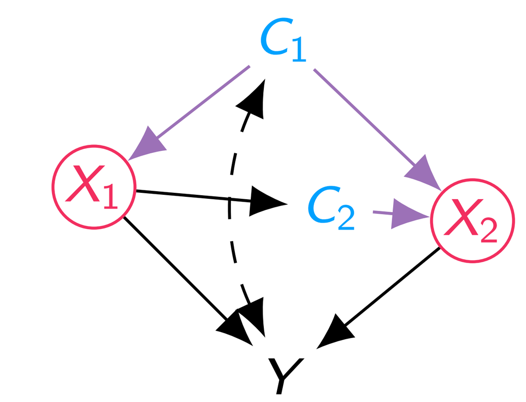
→ the left does *not meet optimality* since the right MPS, when optimized, always yields a higher than or equal expected reward.

- We can show that an agent can optimize policies along the MPSes below out of 15 to **efficiently** and **effectively** explore the environment.

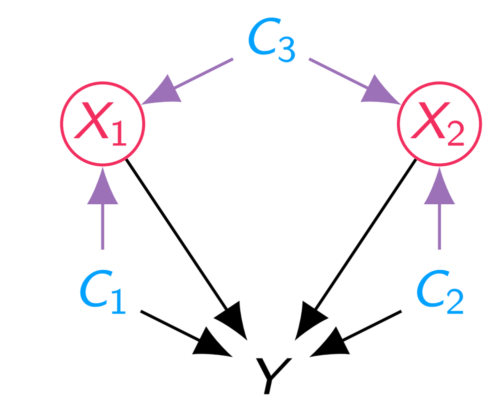


characterization: non-redundancy under optimality

An MPS is **non-redundant** if its actions and contexts are relevant to the reward such that *acting differently* affects the reward. This can be checked by well-known graphical criteria (d-separation & do-calculus).

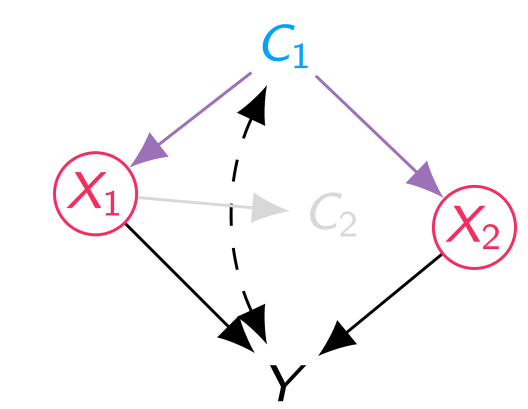


e.g., X_2 listens to C_2 because $C_2 \leftarrow X_1 \rightarrow Y$ conditioned on the other context C_1.

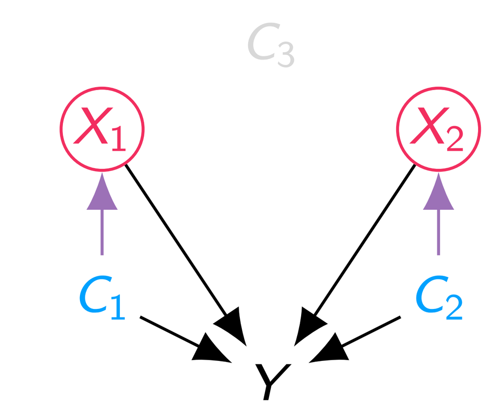


e.g., X_2 listens to C_3 because $C_3 \rightarrow X_1 \rightarrow Y$ conditioned on the other context C_2.

An MPS is **non-redundant under optimality** (NRO) if no MPS subsumed by the MPS can perform equally w.r.t. an optimal expected reward.



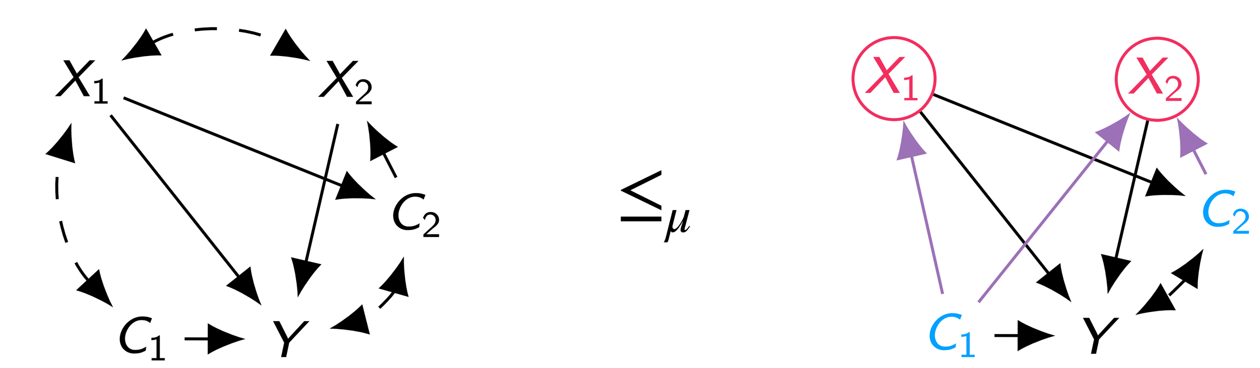
When optimized, X_1 is *deterministic* given C_1. C_2 does not provide any useful information than what C_1 conveys.



The correlation b/w X_1 and X_2 is *irrelevant* to the outcome. They can behave as if C_3 is fixed.

characterization: possible-optimality

An MPS is **possibly-optimal** if the MPS is non-redundant under optimality and there exists no other "better than or equal" MPS w.r.t an optimal expected reward, e.g.,



The left MPS is *not possibly-optimal* because the right MPS can perform always better or equally when optimized.

We provide **necessary conditions** for which an MPS is (i) non-redundant under optimality (NRO) and (ii) possibly-optimal. These help reduce the space of MPSes an agent needs to explore the environment so that it can converge to optimal **efficiently** (NRO) and **effectively** (possible-optimality).