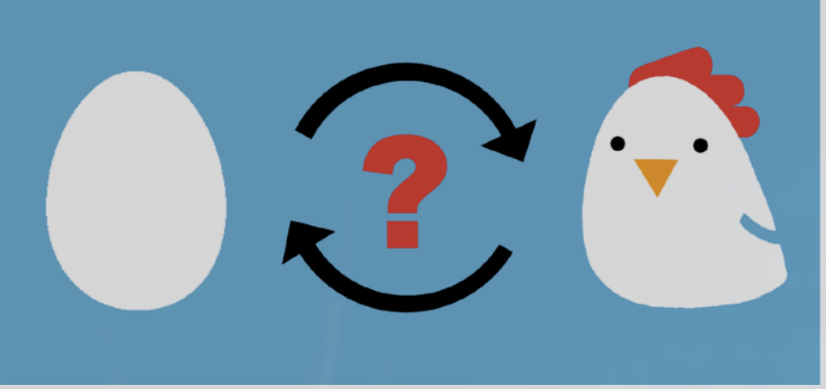


Causality and Machine Learning
(80-816/516)

Classes 21 (April 1, 2025)

Causal Representation Learning 3:
Real Problems to Address

Instructor:

Kun Zhang (kunz1@cmu.edu)

Zoom link: https://cmu.zoom.us/j/8214572323)

Office Hours: W 3:00–4:00PM (on Zoom or in person); other times by appointment

In Addition, Causal Abstraction…

What is 'temperature'?

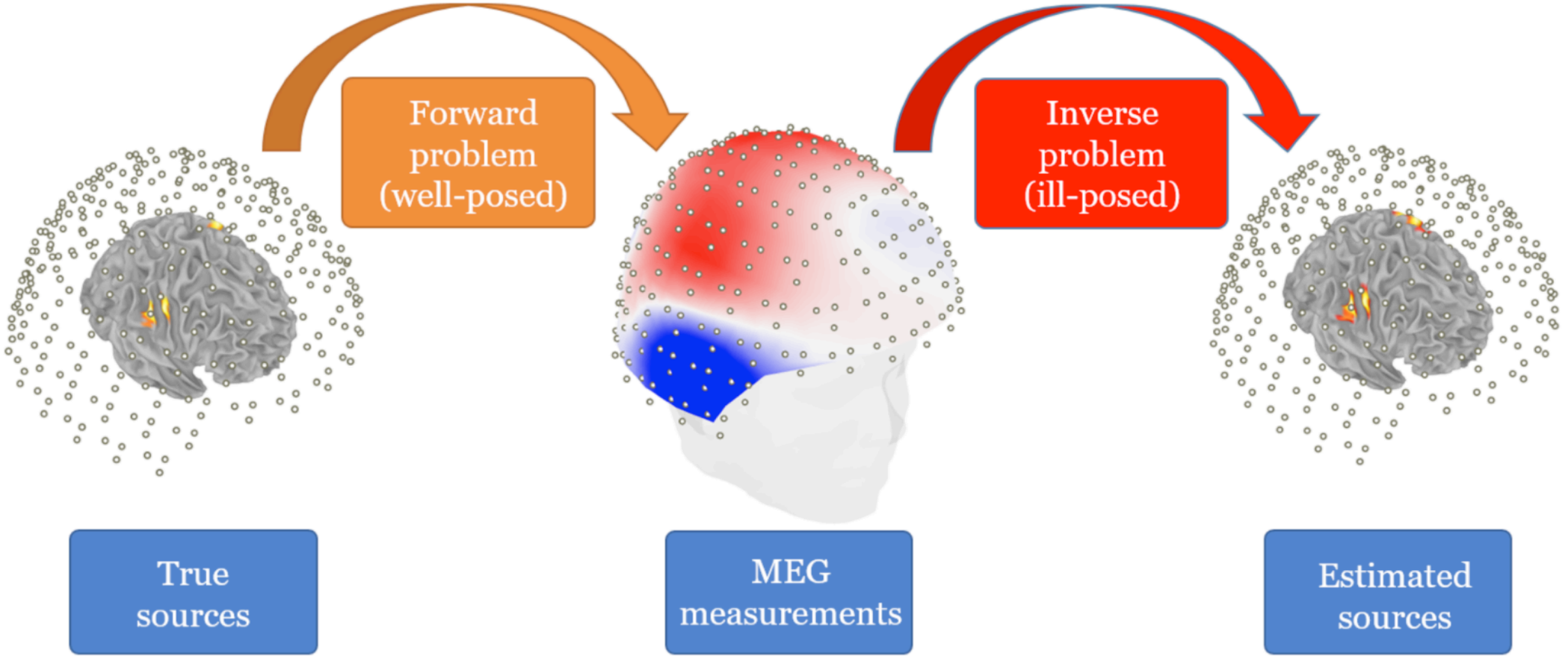
i.i.d. data?	Parametric constraints?	Latent confounders?	What can we get?
Yes	*No*	*No*	*(Different types of) equivalence class*
		Yes	
	Yes	**No**	**Unique identifiability (under structural conditions)**
		Yes	
Non-I, but I.D.	No/Yes	No	(Extended) regression
		Yes	Latent temporal causal processes identifiable!
I., but non-I.D.	No	No	More informative than MEC (CD-NOD)
	Yes		May have unique identifiability
	No	Yes	Changing subspace identifiable
	Yes		Variables in changing relations identifiable

Real Problems Addressed with CRL

- In neuroscience
  - Localization & causal analysis from EEG/MEG data
  - Finding regions of interest from fMRI data
  - Psychometric studies
  - Deep reinforcement learning
  - Multi-model CRL in healthcare
  - Generative AI: Image generation and refinement

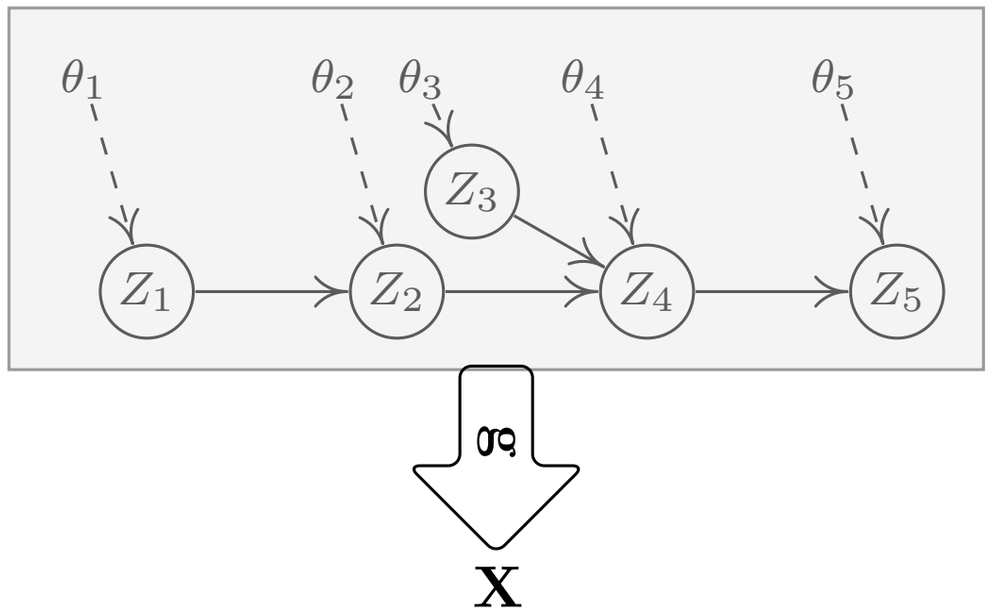
From MEG/EEG

- In both Magnetoencephalography (MEG) and Electroencephalography (EEG), source localization involves inferring the location of brain activity from the measured magnetic or electrical fields
- Which CRL setting/formulation can we use?



Causal Representation Learning from Multiple Distributions: A *General* Setting

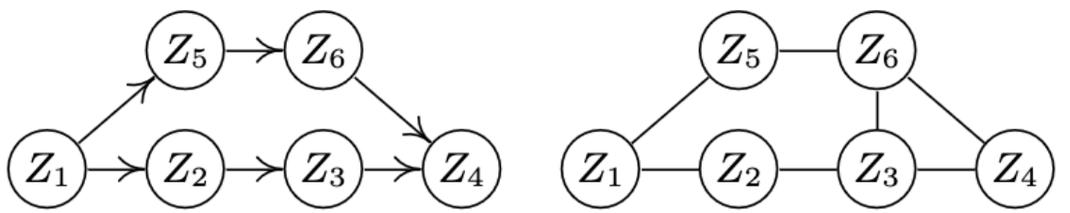
i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes



- Goal: Uncovering hidden variables Z_i with changing causal relations from $\mathbf{X}$ in nonparametric settings

- What is identifiable?
  - Markov network of Z_i
  - Each estimated variable $\tilde{Z}_i$ is a function of Z_i and it **intimate neighbors**

- In this example, each Z_i ($i \neq 4$) can be recovered up to component-wise transformation



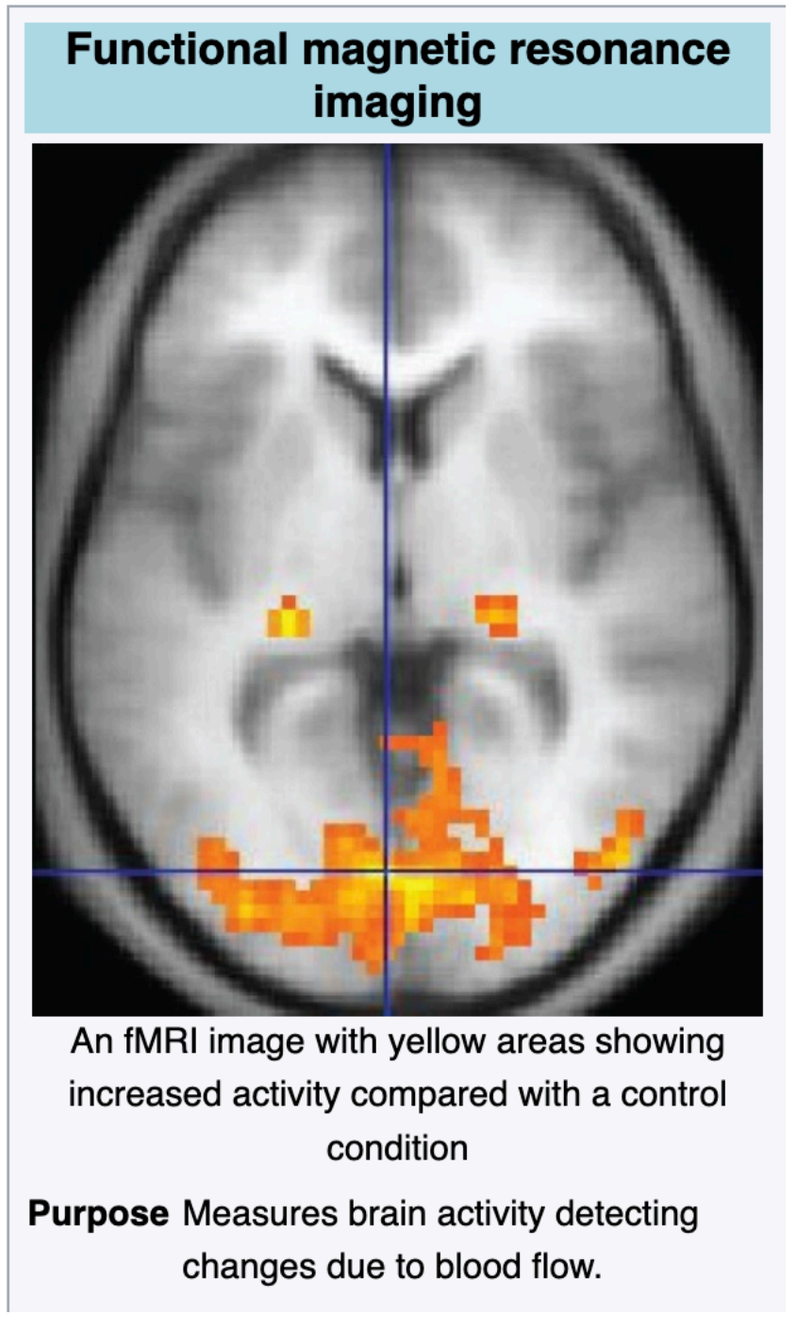
(a) $\mathcal{G}_Z$, the DAG over true latent variables Z_i.

(b) The corresponding Markov network $\mathcal{M}_Z$.

- *Zhang, Xie, Ng, Zheng, "Causal Representation Learning from Multiple Distributions: A General Setting," ICML 2024*

From fMRI

- Functional Magnetic Resonance Imaging (fMRI) measures brain activity by detecting changes associated with blood flow
- The primary form of fMRI uses the blood-oxygen-level dependent (BOLD) contrast
- A voxel is a three-dimensional rectangular cuboid, whose dimensions are set by the slice thickness, the area of a slice, and the grid imposed on the slice by the scanning process.
- Voxel data can be very noisy. Going to regions of interest?



https://en.wikipedia.org/wiki/Functional_magnetic_resonance_imaging

Abstracting Causal Models

Sander Beckers
Dept. of Philosophy and Religious Studies
Utrecht University
Utrecht, Netherlands
srekcebrednas@gmail.com

Joseph Y. Halpern
Dept. of Computer Science
Cornell University
Ithaca, NY 14853
halpern@cs.cornell.edu

Abstract

We consider a sequence of successively more restrictive definitions of abstraction for causal models, starting with a notion introduced by Rubenstein et al. (2017) called *exact transformation* that applies to probabilistic causal models, mov
a notion of *uniform transformation* that applies to deter
tic causal models and does not allow differences to be h
by the "right" choice of distribution, and then to *abstra*
where the interventions of interest are determined by th
from low-level states to high-level states, and *strong ab*
tion, which takes more seriously all potential interventi
a model, not just the allowed interventions. We show tha
cedures for combining micro-variables into macro-var
are instances of our notion of strong abstraction, as are
examples considered by Rubenstein et al.

1 Introduction

Rubenstein et al. (2017) (RW$^+$ from now on) provided an arguably more general approach to abstraction. They defined a notion of an *exact transformation* between two causal models. They suggest that if there is an exact transformation τ from causal model M_1 to M_2, then we should think of M_2

What is 'temperature'?

Temperature as a Measure of Motion: Temperature reflects the average speed and motion of particles within a substance.

Be Aware of Causal Abstraction

- The Mpemba effect is a **counterintuitive** phenomenon where, under certain conditions, initially hot water can freeze faster than initially cold water.
- "When a warm sample of water is placed in a cold environment, the part of it next to the walls of the container gets cooled quickly while the inner part remains its temperature. A temperature gradient is thereby induced inside of the sample which causes convective heat transport. The greater heat gradient gets, the convection is more expressed, and the overall cooling of the sample is faster, since the heat gradient on the container walls is maintained."

Mpemba effect from a viewpoint of an experimental physical chemist

by Nikola Bregović

Real Problems Addressed with CRL

- In neuroscience
  - Localization & causal analysis from EEG/MEG data
  - Finding regions of interest from fMRI data
- Psychometric studies
- Deep reinforcement learning
- Multi-model CRL in healthcare
- Generative AI: Image generation and refinement

A Problem in Psychology: Finding Underlying Mental Conditions?

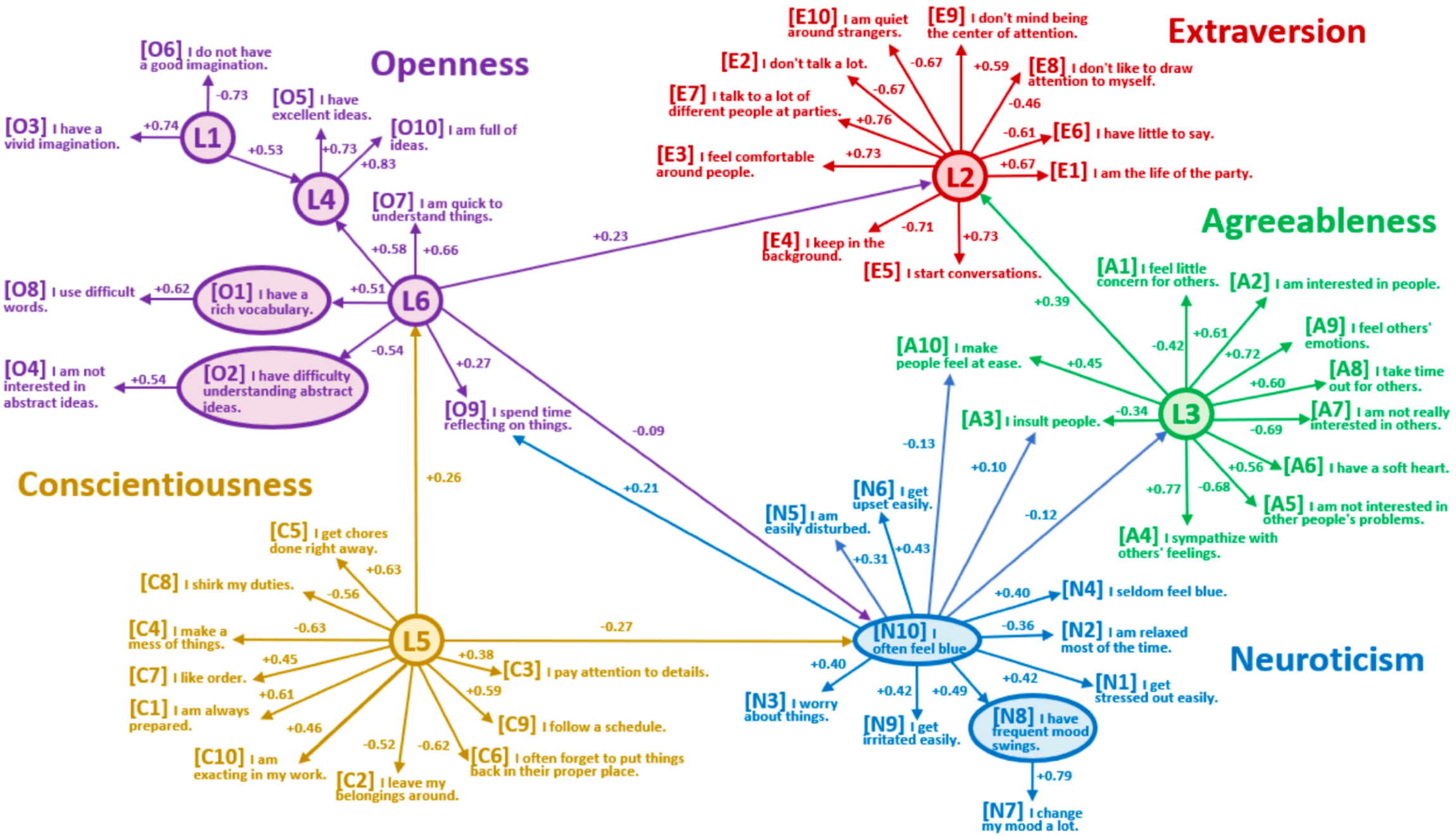
i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- 50 questions for big 5 personality test

race	age	engnat	gender	hand	source	country	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	A1	A2	A3	A4	A5
3	53	1	1	1	1	US	4	2	5	2	5	1	4	3	5	1	1	5	2	5	1	1	1	1	1	1	1	5	1	5	2
13	46	1	2	1	1	US	2	2	3	3	3	3	1	5	1	5	2	3	4	2	3	4	3	2	2	4	1	3	3	4	4
1	14	2	2	1	1	PK	5	1	1	4	5	1	1	5	5	1	5	1	5	5	5	5	5	5	5	5	5	1	5	5	1
3	19	2	2	1	1	RO	2	5	2	4	3	4	3	4	4	5	5	4	4	2	4	5	5	5	4	5	2	5	4	4	3
11	25	2	2	1	2	US	3	1	3	3	3	1	3	1	3	5	3	3	3	4	3	3	3	3	3	4	5	5	3	5	1
13	31	1	2	1	2	US	1	5	2	4	1	3	2	4	1	5	1	5	4	5	1	4	4	1	5	2	2	2	3	4	3
5	20	1	2	1	5	US	5	1	5	1	5	1	5	4	4	1	2	4	2	4	2	2	3	2	2	2	5	5	1	5	1
4	23	2	1	1	2	IN	4	3	5	3	5	1	4	3	4	3	1	4	4	4	1	1	1	1	1	1	2	5	1	4	3
5	39	1	2	3	4	US	3	1	5	1	5	1	5	2	5	3	2	4	5	3	3	5	5	4	3	3	1	5	1	5	1
3	18	1	2	1	5	US	1	4	2	5	2	4	1	4	1	5	5	2	5	2	3	4	3	2	3	4	2	3	1	4	2
3	17	2	2	1	1	IT	1	5	2	5	1	4	1	4	1	5	5	3	5	3	2	5	3	3	4	3	2	4	2	4	1
13	15	2	1	1	1	IN	3	3	5	3	3	3	2	4	3	3	1	5	3	3	2	3	2	3	2	4	4	4	2	2	5
13	22	1	2	1	2	US	3	3	4	2	4	2	2	3	4	3	3	3	3	3	2	2	4	4	2	3	1	4	1	5	1
3	21	1	2	1	5	US	1	3	2	5	1	1	1	5	1	5	5	3	5	2	5	5	3	2	5	3	1	1	1	4	2
3	28	2	2	1	2	US	3	3	3	4	3	2	2	4	3	5	2	4	4	4	4	4	2	2	3	2	1	4	2	4	2
3	21	1	1	1	5	US	2	3	2	3	3	1	1	3	4	4	2	4	2	4	1	2	2	2	2	2	4	2	4	2	5
13	19	1	2	1	2	FR	1	3	2	4	2	4	1	4	3	4	4	2	3	2	1	3	1	2	2	3	4	2	3	1	4
3	21	1	2	1	5	US	4	1	5	2	5	1	5	3	5	1	5	2	5	2	3	3	3	3	4	2	1	5	2	5	2

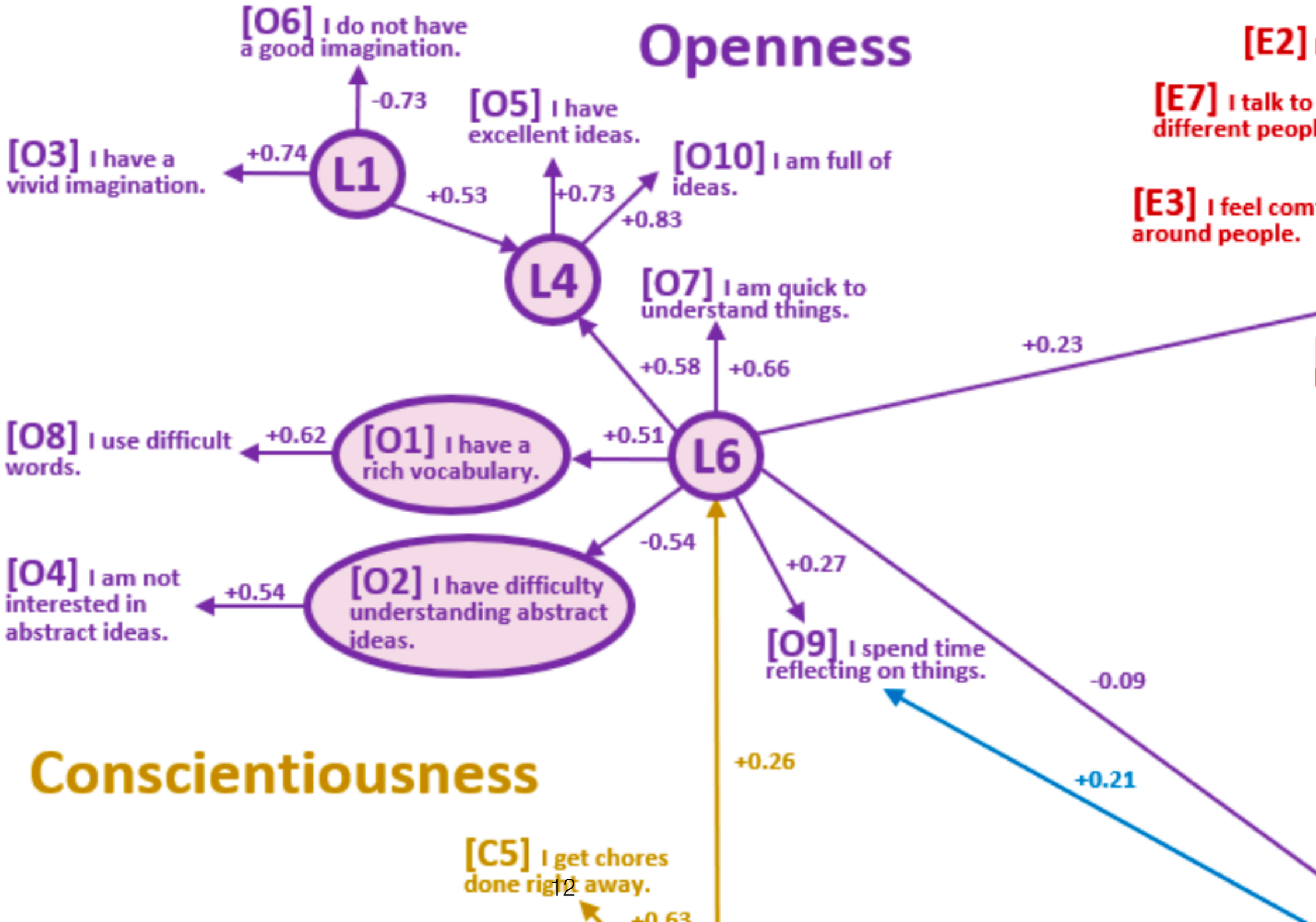
Example: Big 5 Questions Are Well Designed but…

Big 5: openness; conscientiousness; extraversion; agreeableness; neuroticism



- Dong, Huang, Ng, Song, Zheng, Jin, Legaspi, Spirtes, Zhang, "A Versatile Causal Discovery Framework to Allow Causally-Related Hidden Variables," ICLR 2024

Example: Big 5 Questions Are Well Designed but…



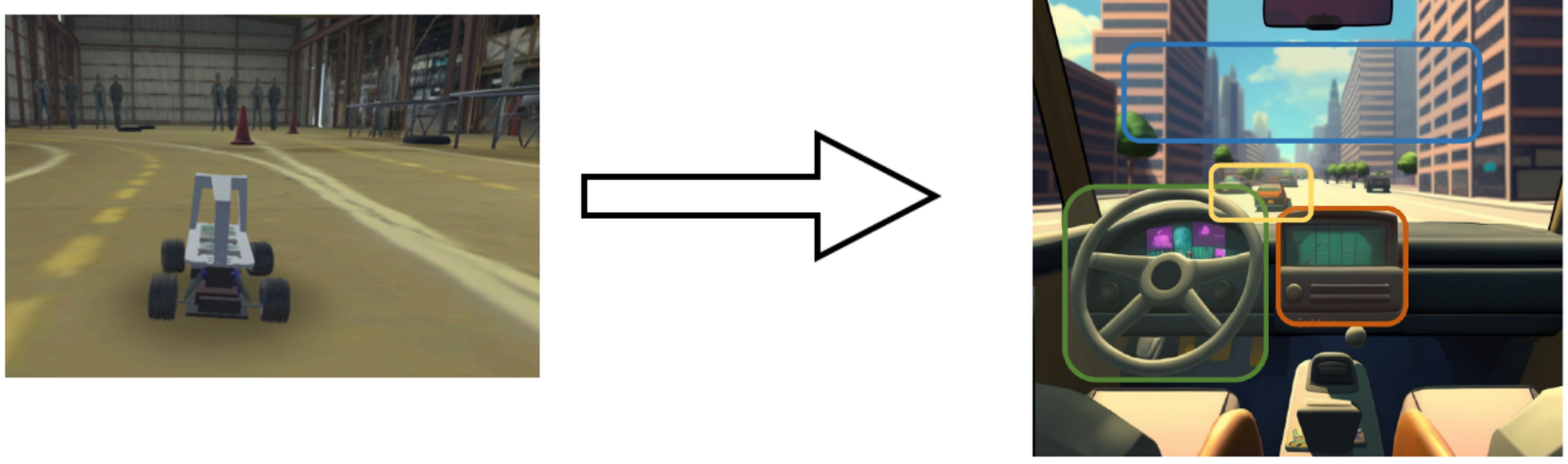
- *Dong, Huan*

Real Problems Addressed with CRL

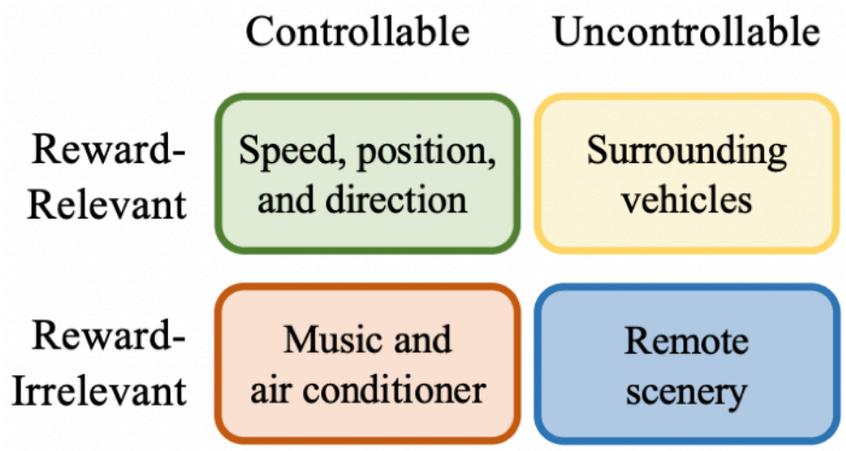
- In neuroscience
  - Localization & causal analysis from EEG/MEG data
  - Finding regions of interest from fMRI data
  - Psychometric studies
- Deep reinforcement learning
- Multi-model CRL in healthcare
- Generative AI: Image generation and refinement

A Causal Perspective on Reinforcement Learning

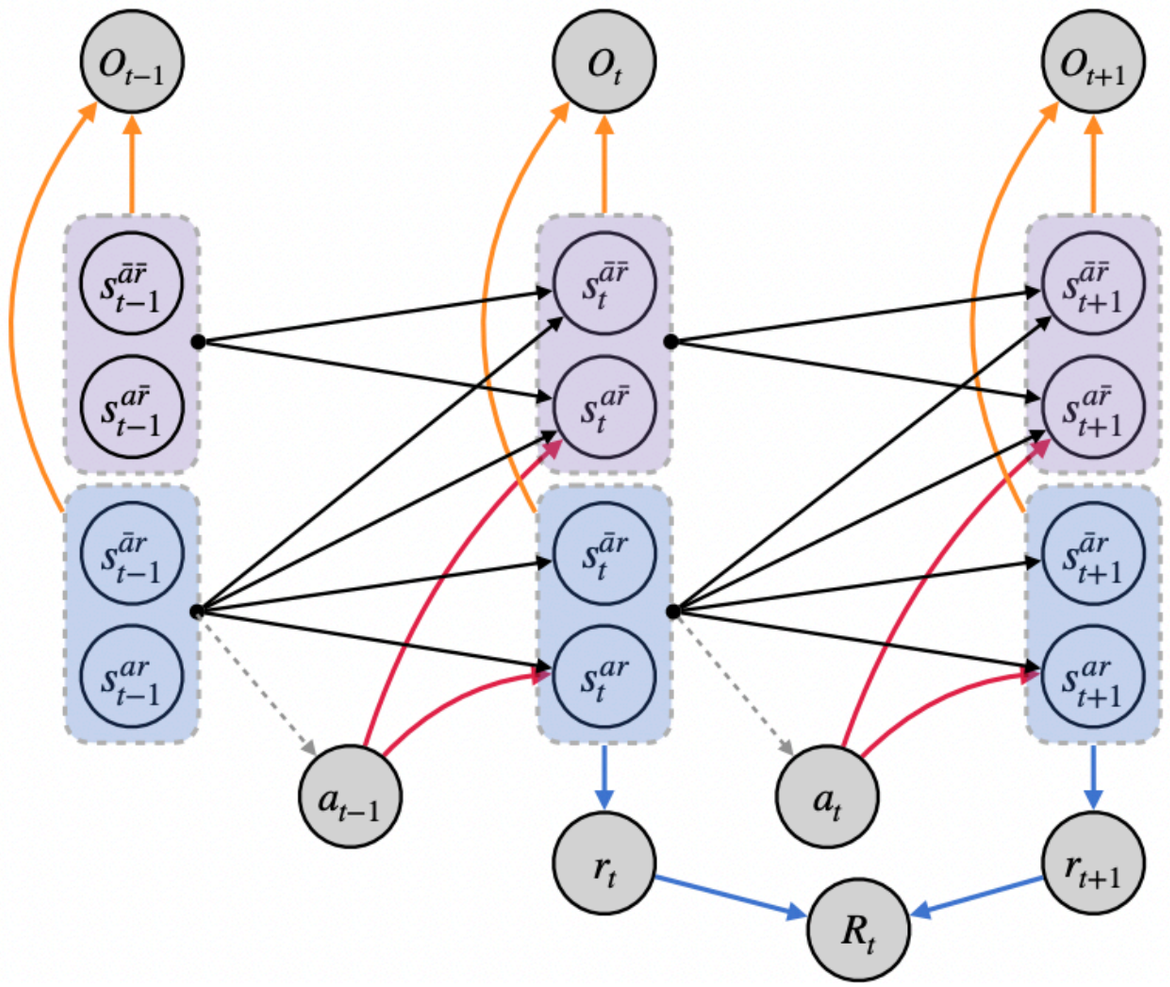
- Potential issues in deep RL algorithms
  - Lack interpretability
  - Not generalize well
  - Data hungry
- Mitigate such issues through causal representations and graph structures



Four Categories of State Representations in RL

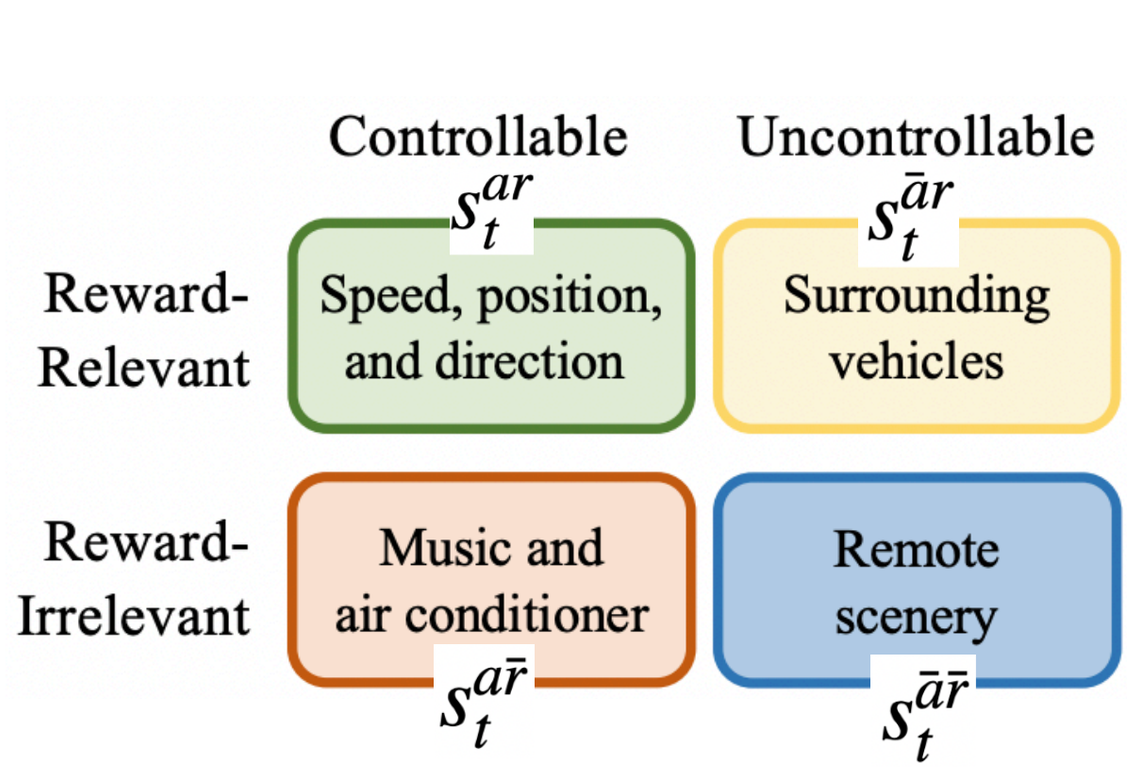


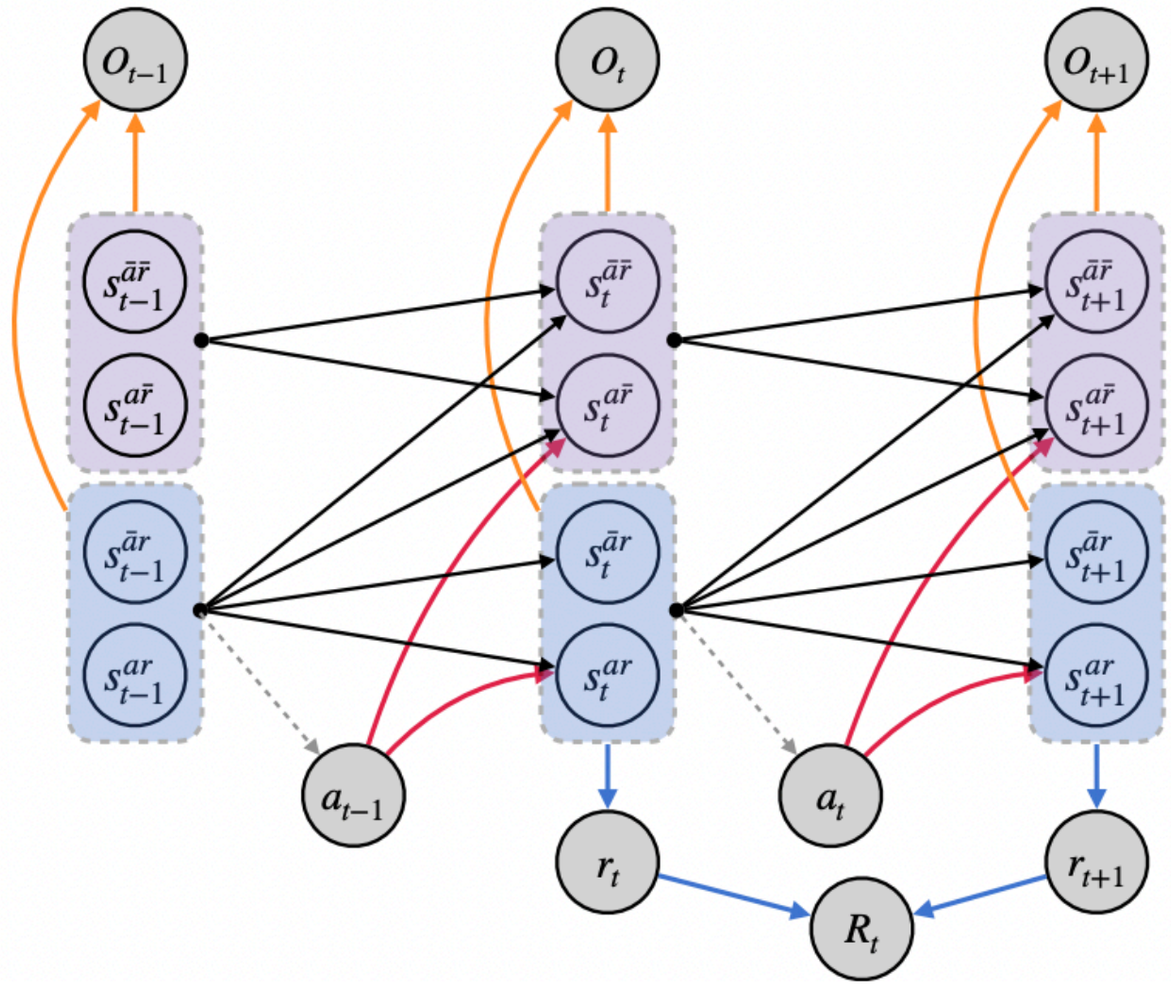
- s_t^{ar} : controllable and reward-relevant state variables
- $s_t^{\bar{a}r}$: reward-relevant state variables that are beyond our control
- $s_t^{a\bar{r}}$: controllable but reward-irrelevant factors
- $s_t^{\bar{a}\bar{r}}$: uncontrollable and reward-irrelevant latent variables



Liu*, Huang*, Zhu, Tian, Gong, Yu, Zhang. Learning world models with identifiable factorization. Arxiv, 2023.

Four Categories of State Representations in RL

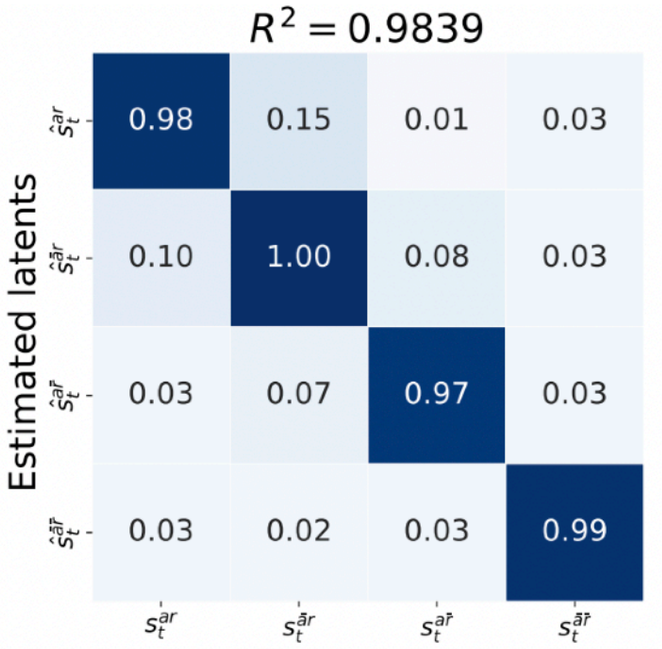




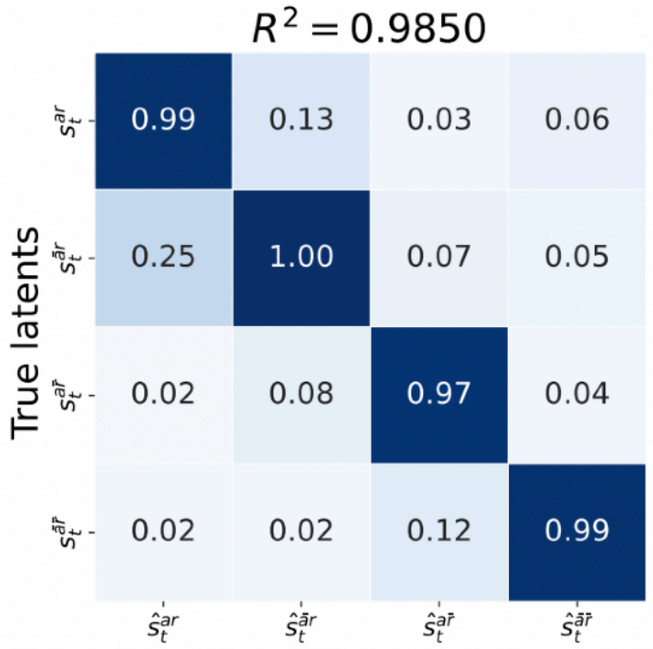
- Each category is identifiable!

- Liu*, Huang*, Zhu, Tian, Gong, Yu, Zhang. Learning world models with identifiable factorization. NeurIPS 2023

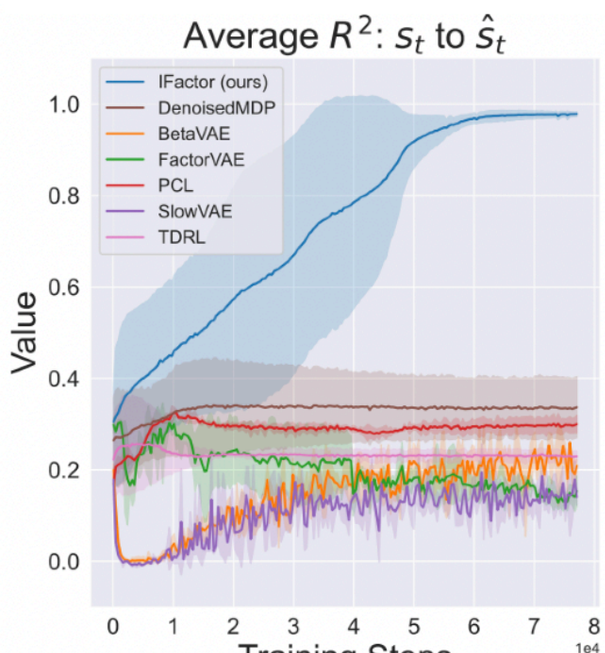
Experimental Results on Latent States Recovery



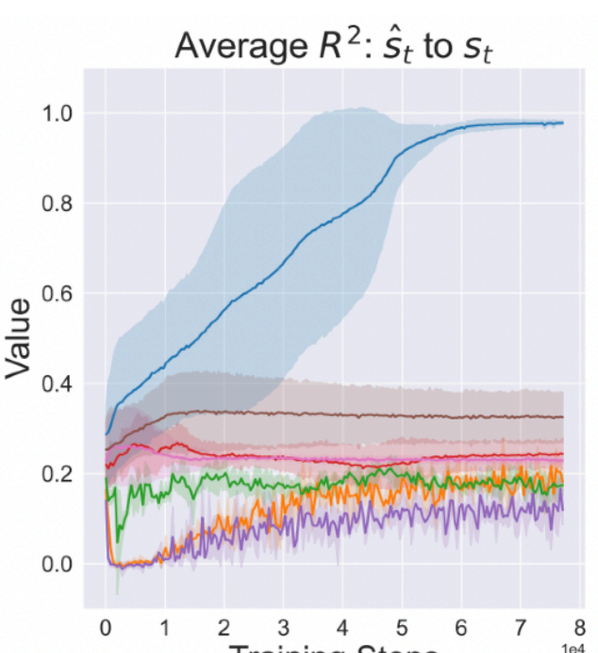
(a)



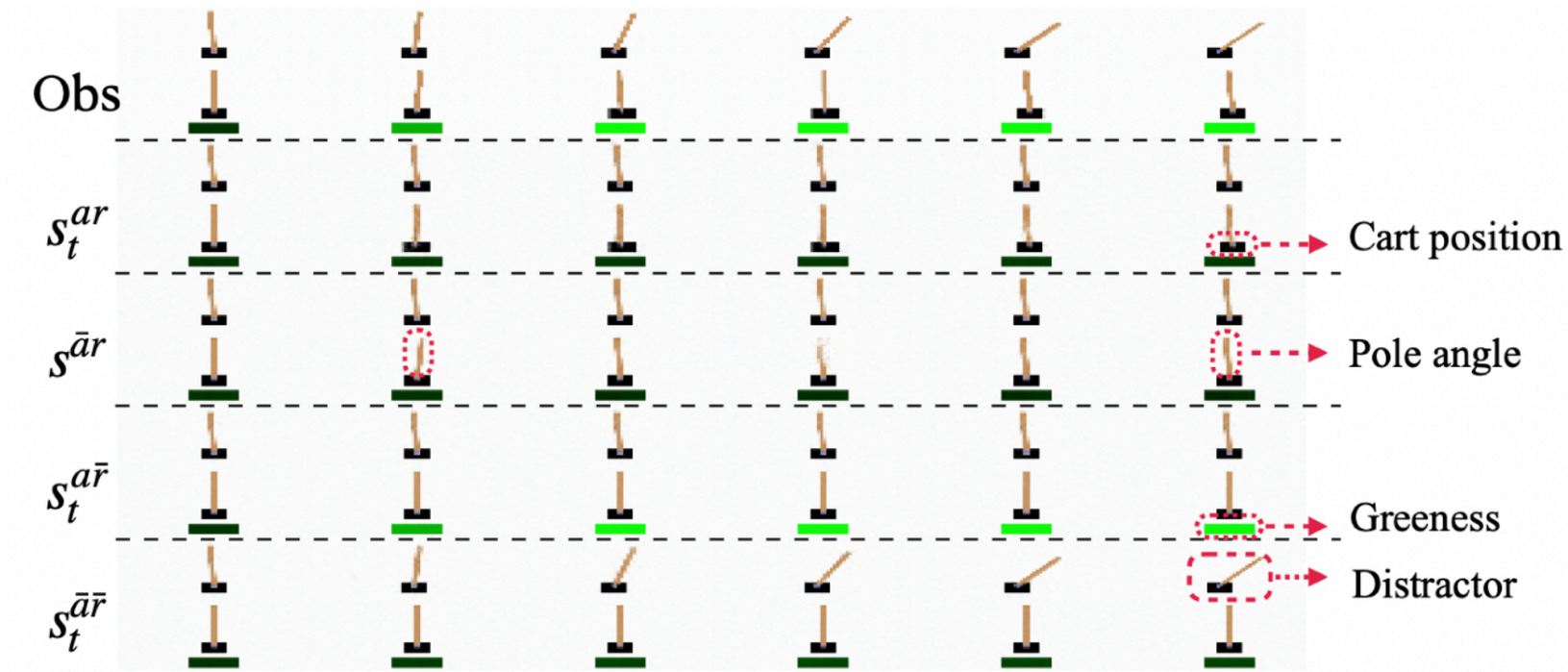
(b)



(c)

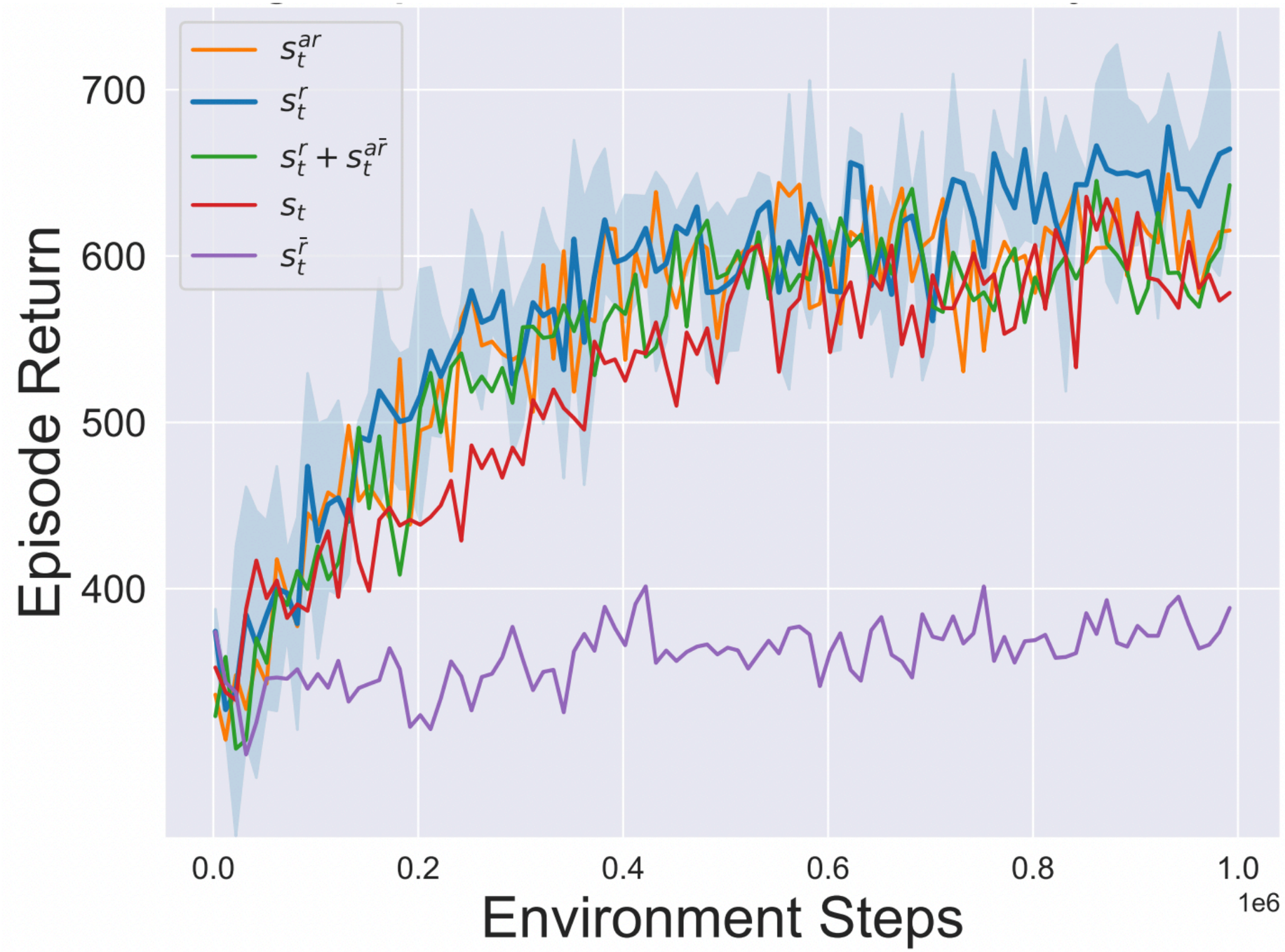


(d)



Modified Cartpole with two distractors

Experimental Results on Policy Learning



Episode return with different state representations

Real Problems Addressed with CRL

- In neuroscience
  - Localization & causal analysis from EEG/MEG data
  - Finding regions of interest from fMRI data
  - Psychometric studies
  - Deep reinforcement learning
- Multi-model CRL in healthcare
- Generative AI: Image generation and refinement

CAUSAL REPRESENTATION LEARNING FROM MULTIMODAL BIOMEDICAL OBSERVATIONS

Yuewen Sun[1,2]*, **Lingjing Kong**[2]*, **Guangyi Chen**[1,2], **Loka Li**[1], **Gongxu Luo**[1], **Zijian Li**[1], **Yixuan Zhang**[1], **Yujia Zheng**[2], **Mengyue Yang**[3], **Petar Stojanov**[4], **Eran Segal**[1], **Eric P. Xing**[1,2], **Kun Zhang**[1,2]
[1]Mohamed bin Zayed University of Artificial Intelligence, [2]Carnegie Mellon University, [3]University of Bristol, [4]Broad Institute of MIT and Harvard

ABSTRACT

Prevalent in biomedical applications (e.g., human phenotype research), multimodal datasets can provide valuable insights into the underlying physiological mechanisms. However, current machine learning (ML) models designed to analyze these datasets often lack interpretability and identifiability guarantees, which are essential for biomedical research. Recent advances in causal representation learning have shown promise in identifying interpretable latent causal variables with formal theoretical guarantees. Unfortunately, most current work on multimodal distributions either relies on restrictive parametric assumptions or yields only coarse identification results, limiting their applicability to biomedical research that favors a detailed understanding of the mechanisms.

In this work, we aim to develop flexible identification conditions for multimodal data and principled methods to facilitate the understanding of biomedical datasets. Theoretically, we consider a nonparametric latent distribution (c.f., parametric assumptions in previous work) that allows for causal relationships across potentially different modalities. We establish identifiability guarantees for each latent component, extending the subspace identification results from previous work. Our key theoretical contribution is the structural sparsity of causal connections between modalities, which, as we will discuss, is natural for a large collection of biomedical systems. Empirically, we present a practical framework to instantiate our theoretical insights. We demonstrate the effectiveness of our approach through

CRL from Multi-Modal Data?

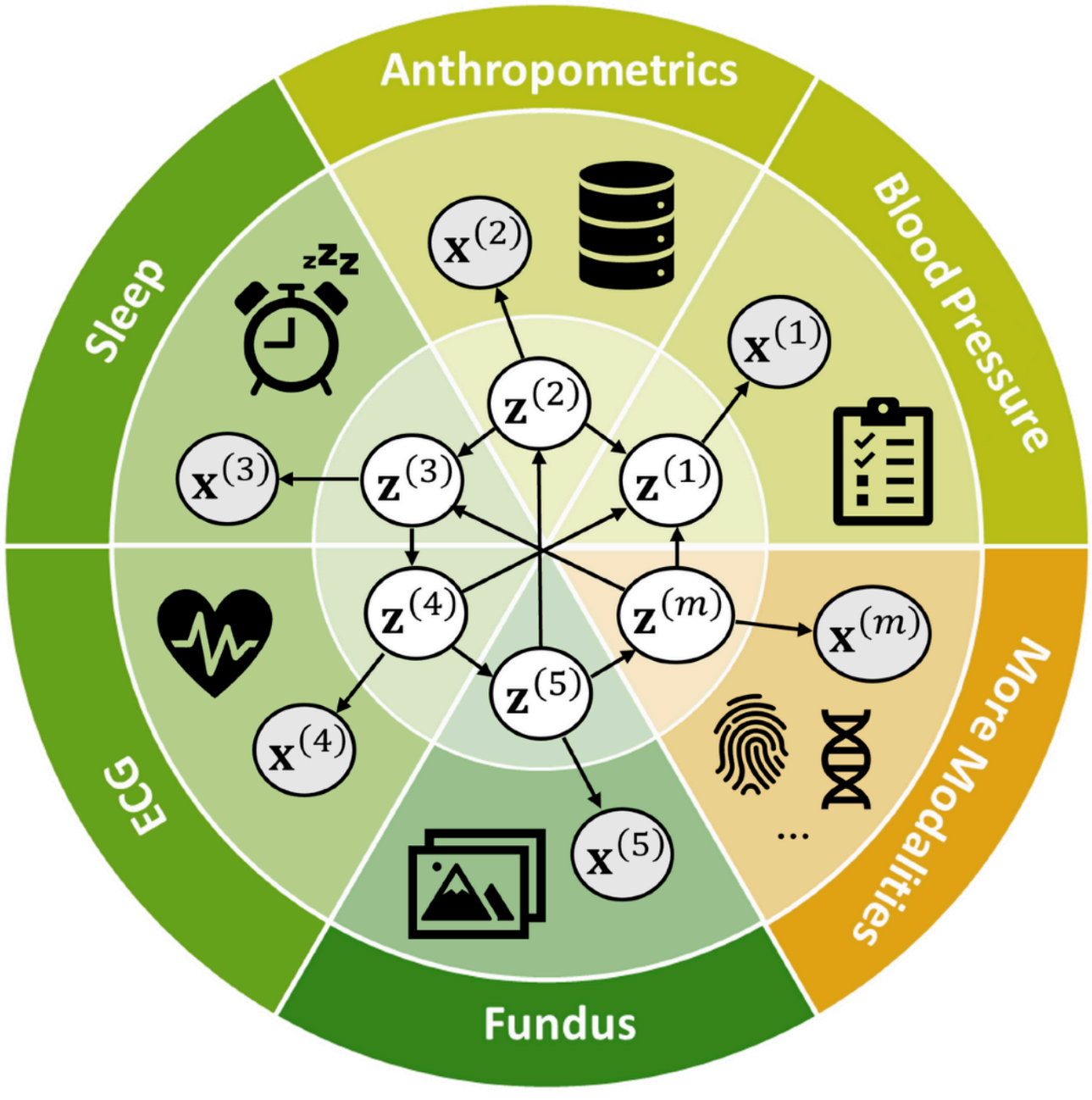


Figure 1: Multimodal data with causal latent variables.

Part of the Result on Human Phenotype Data

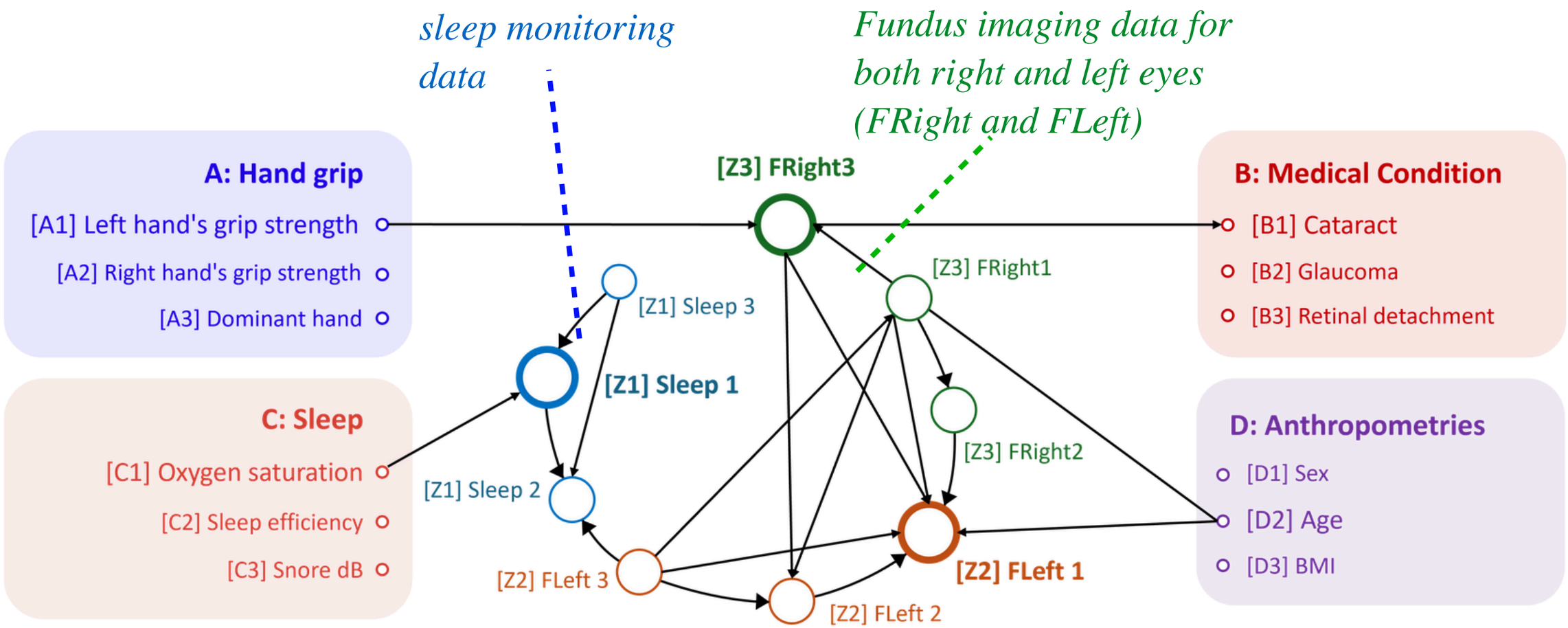


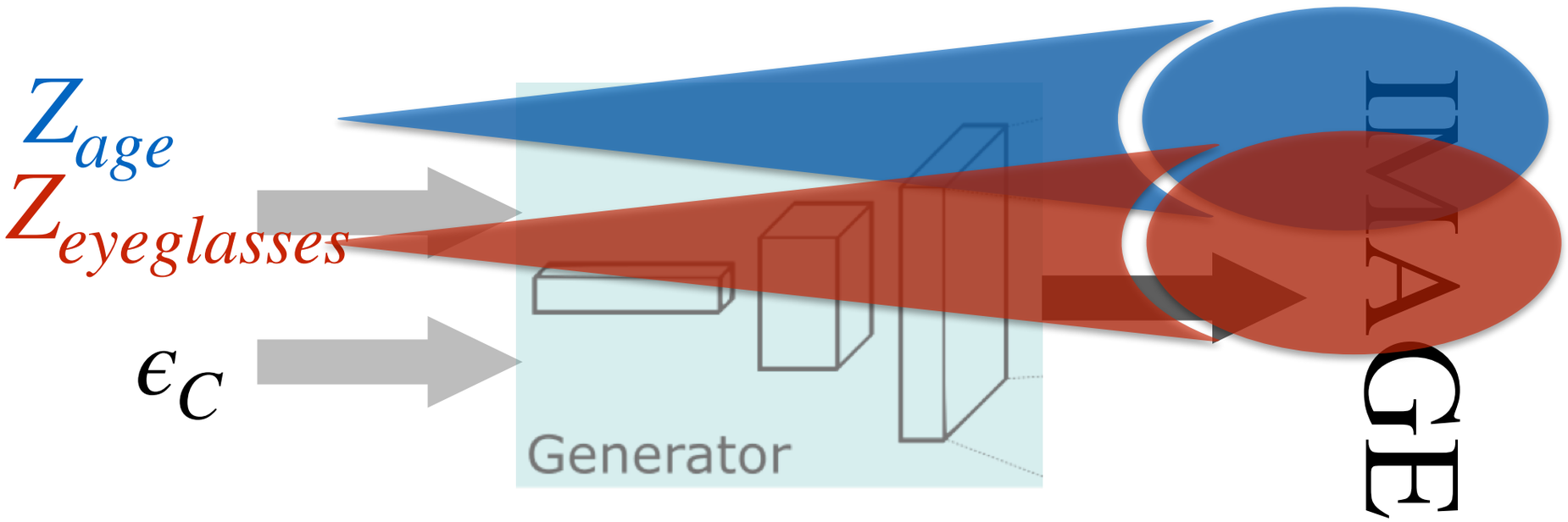
Figure 6: Causal analysis results across different modalities, including hand grip, medical conditions, sleep, and anthropometries. We ran the causal algorithm on all variables but reported only the causal relations that have direct connections to the estimated latent variables for clarity.

Real Problems Addressed with CRL

- In neuroscience
  - Localization & causal analysis from EEG/MEG data
  - Finding regions of interest from fMRI data
  - Psychometric studies
  - Deep reinforcement learning
  - Multi-model CRL in healthcare
- Generative AI: Image generation and refinement

Dealing with Age vs. Eyeglasses

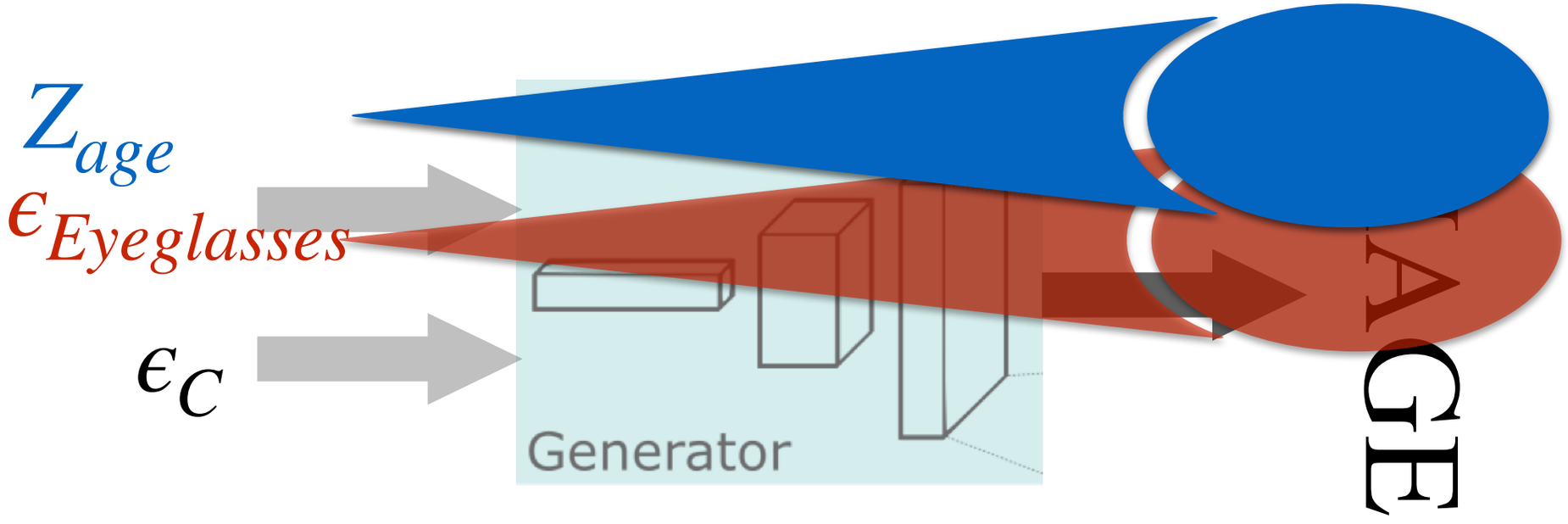
- They are dependent in the data
- What if we treat them as features that we can manipulate independently/separately?



- So, changing one of them may lead to change in the perception of the other

Causal Asymmetry

- Age → Eyeglasses: interventions on Age may change Eyeglasses, but not the other way around
- Functional causal model: Eyeglasses = f(Age, $\epsilon_{Eyeglasses}$), where $\epsilon_{Eyeglasses} \perp\!\!\!\perp$ Age
- Intervention on Eyeglasses via changing $\epsilon_{Eyeglasses}$!

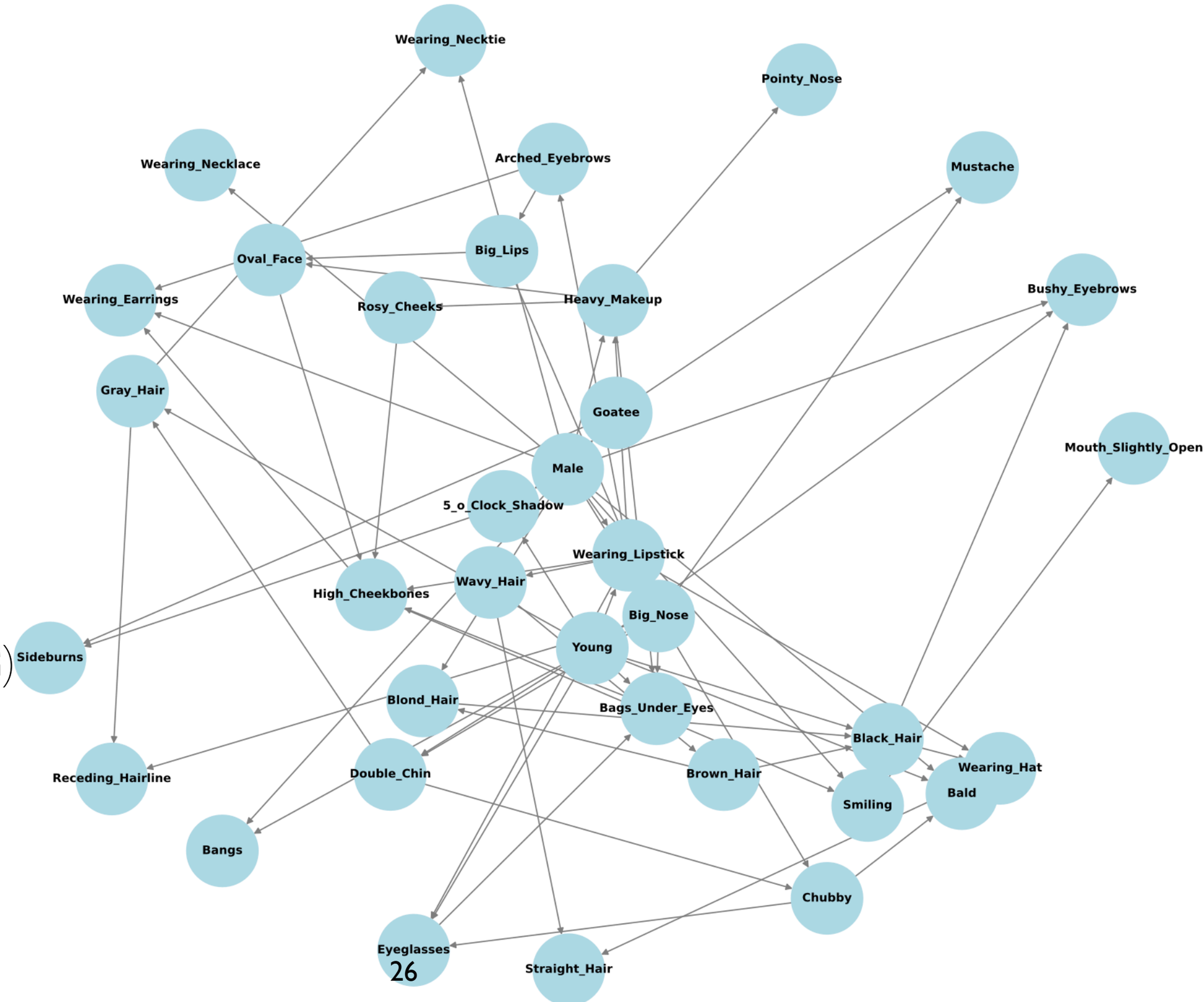


- Moreover, minimal changes for both Z_{age} and $\epsilon_{Eyeglasses}$

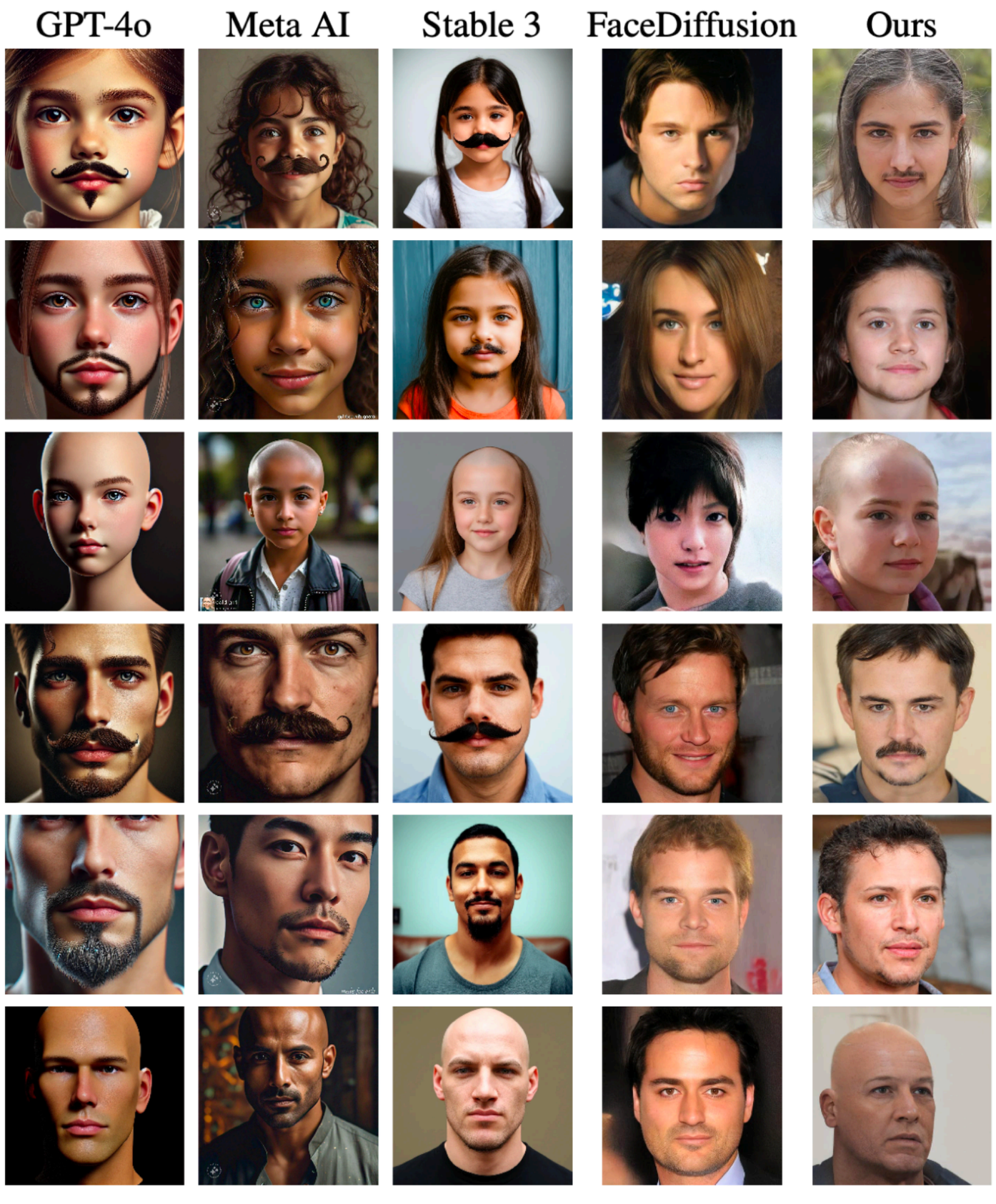
- *Xie, Zheng, and Zhang, under submission*

Causal Graph Among Labels in the Data

- FFHQ dataset (Karras et al., 2019)
- Pre-trained classifier to obtain 37 attributes
- Causal graph learned by causal-learn (PC)
- Then perform image generation or editing



Comparisons: Generation with Different Conditions



Corresponding to:

girl with mustache,

girl with goatee,

bald girl,

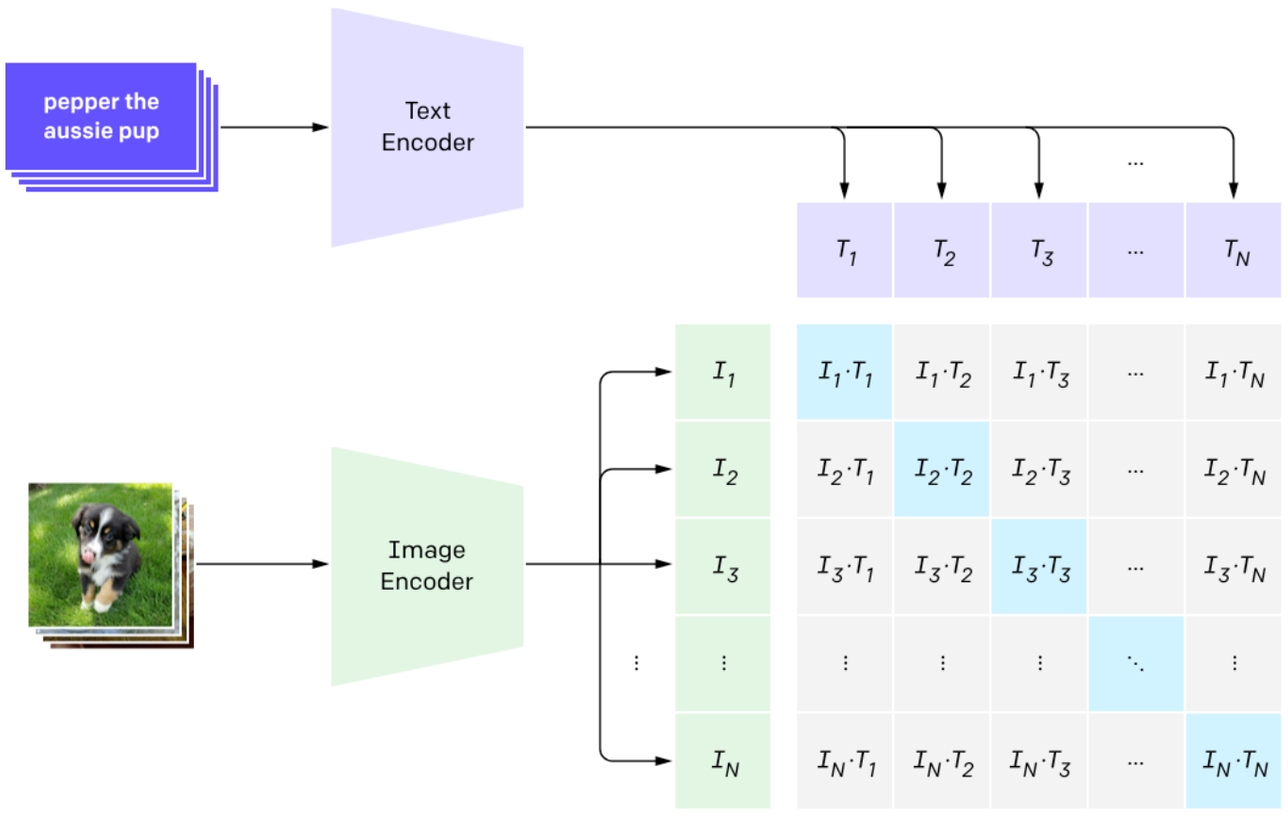
male with mustache,

male with goatee, and

bald male

CLIP doesn't Have a Generative View

- Going beyond CLIP (Contrastive Language-Image Pretraining) model
- We developed SmartCLIP to deal with *missing text* info and *unpaired data*
  - Better alignment
- Causal/generative view?



- *Xie, Kong, Zheng, Yao, Tang, Xing, Chen, and Zhang, under submission*

Motivation: Controllability for Image Generation / Editing

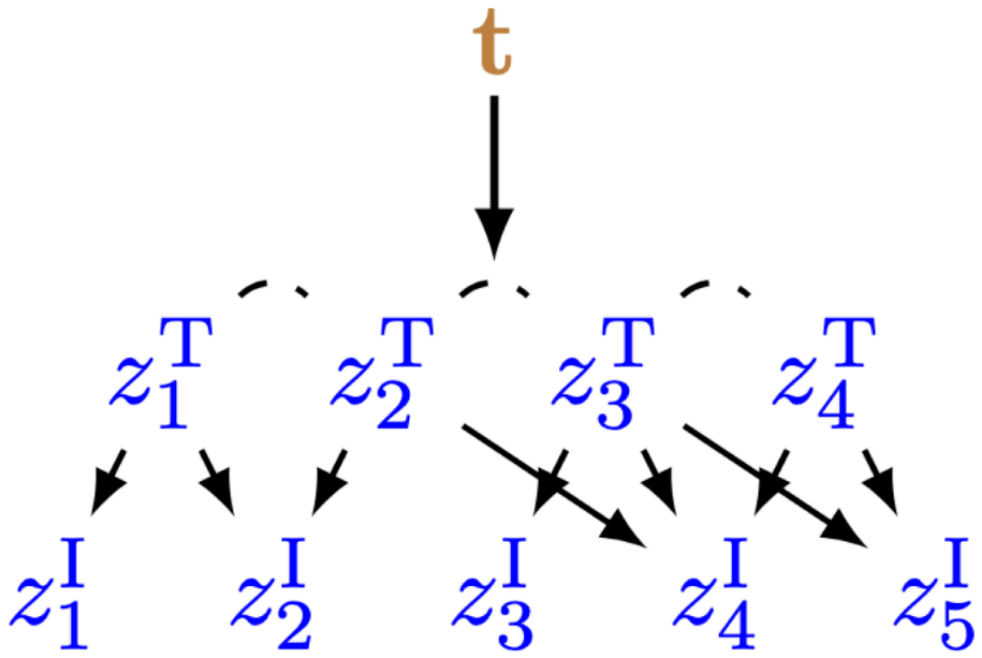
- Existing text-to-image (T2I) models are not controllable: editing a specific feature through text often causes unwanted changes
- Example:

“Happy”

“Angry”

“Surprised”

From Text to Images: The Process



t: text

z_i^{T}: atomic textual concepts

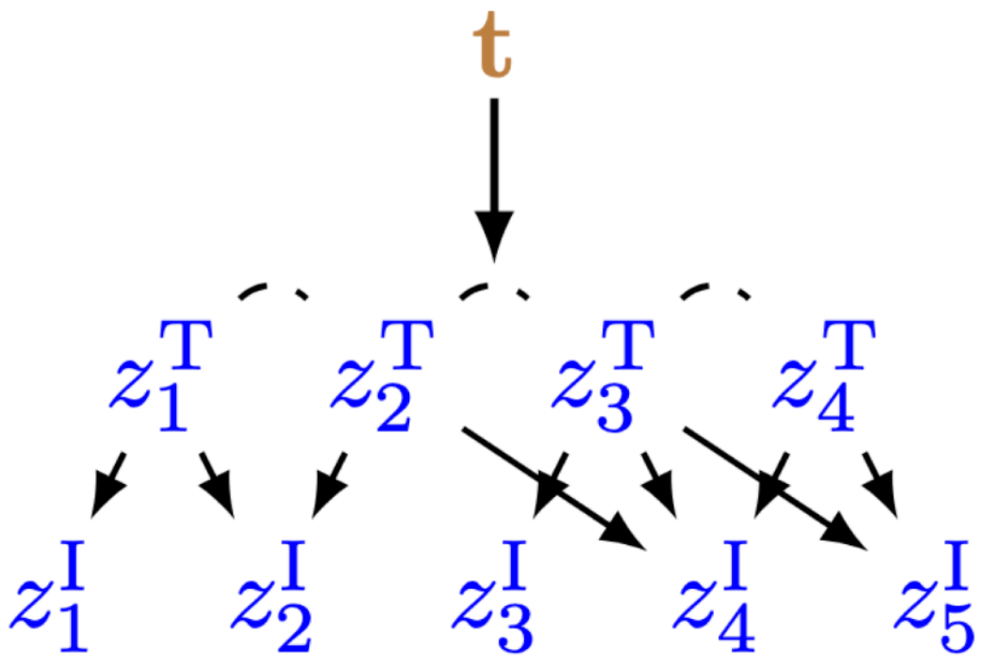
z_j^{I}: atomic visual concepts

i: images

- Text and images have atomic concepts
- Textual atomic concepts determine their visual counterparts: why?

– *Xie, Kong, Zheng, Tang, Xing, Chen, and Zhang, under submission*

Learning Identifiable Concepts for Controllability



t: text

z_i^{T}: atomic textual concepts

z_j^{I}: atomic visual concepts

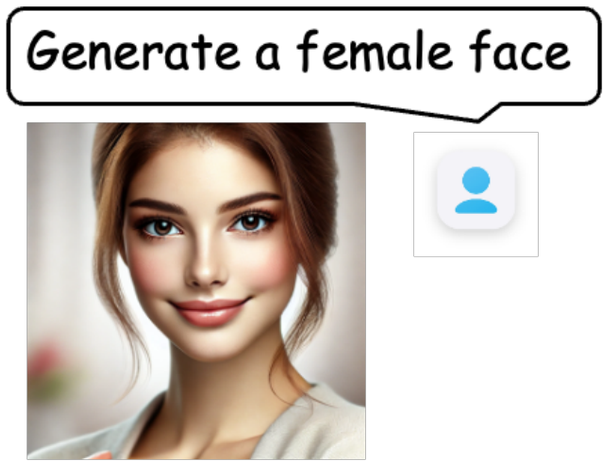
i: images

Certain sparsity constraints on the cross links + conditional independence of image concepts $\Rightarrow$ **identifiable** concepts:

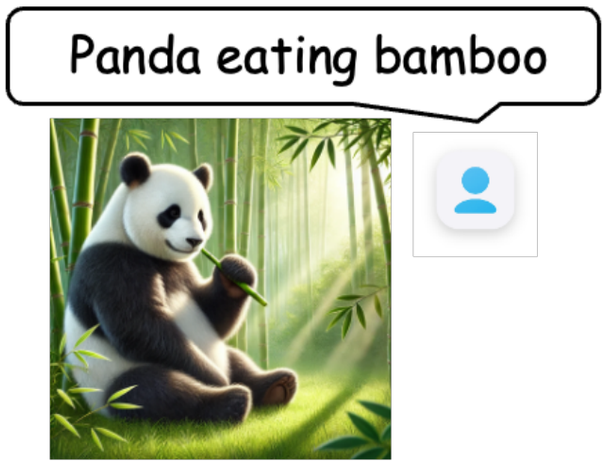
1. Learning *disentangled*, *atomic* concepts z_m^{T} and z_n^{I}.

2. *Aligning* them.

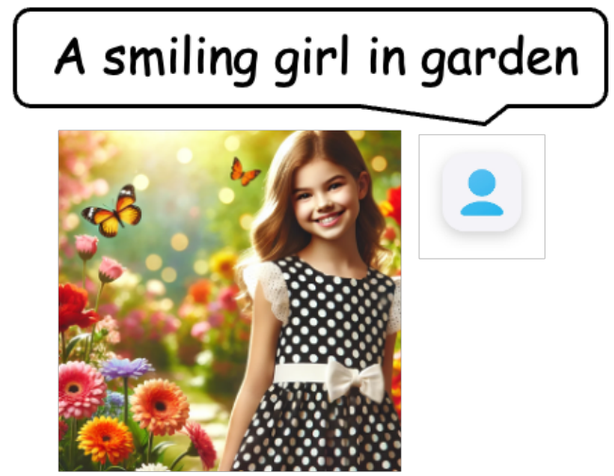
Prevailing generative AI tools: not controllable



Example1: Add Mustache



Example2: Change the style

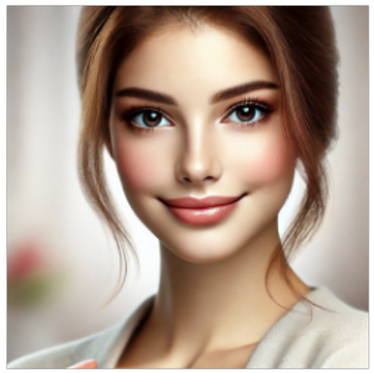


Example3: Change facial expression

Our Causal GenAI Enables Precise Control & Refinement

Example1: Add Mustache

Generate a female face



OpenAI

Can you add mustache ?

Example 3: Change facial expression

A smiling girl in garden

OpenAI

She looks surprised

Our Causal GenAI Enables Precise Control & Refinement

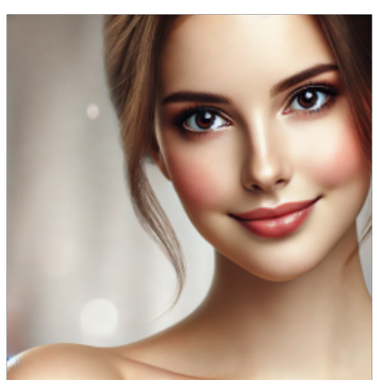
Example1: Add Mustache

Generate a female face

OpenAI

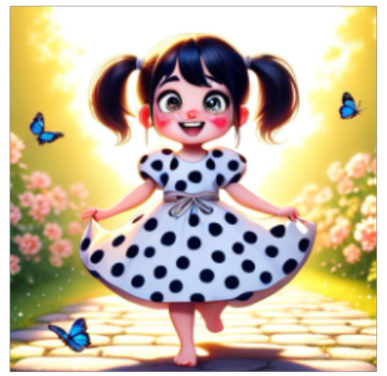
Can you add mustache ?

Ours



Example 3: Change facial expression

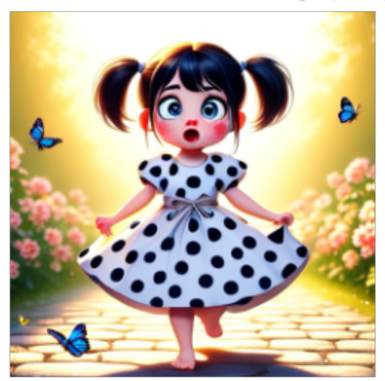
A smiling girl in garden



OpenAI

She looks surprised

Ours



Summary: Real Problems of CRL

- In neuroscience
  - Localization & causal analysis from EEG/MEG data (latent variables; changing distributions)
  - Finding regions of interest from fMRI data (causal abstraction)
- Psychometric studies (latent variables, i.i.d. case)
- Deep reinforcement learning (temporal constraints)
- Multi-model CRL in healthcare (multi-modal learning)
- Generative AI: Image generation and refinement (latent variables; changing distributions)