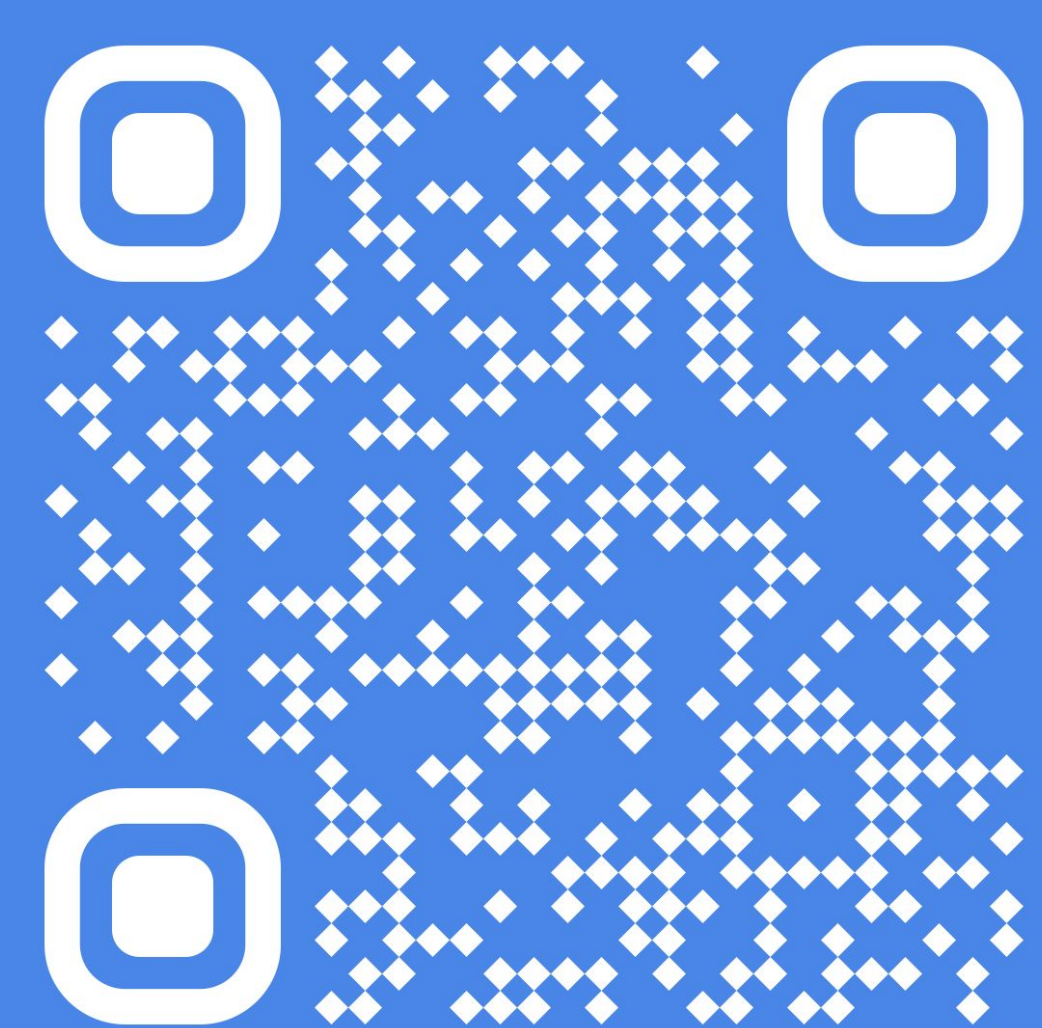




CAUSAL CONCEPT GRAPH MODELS: BEYOND CAUSAL OPACITY IN DEEP LEARNING

Gabriele Dominici, Pietro Barbiero, Mateo Espinosa Zarlenga, Alberto Termine, Martin Gjoreski, Giuseppe Marra, Marc Langheinrich*

*Concept Graph Models enable **understanding** & **intervening** in a model's **causal reasoning process**.*

Causal Opacity

Causal Opacity: lack of transparency in a model's internal causal reasoning

Causal Discovery: recover the causal structure of the real world.

Causal Opacity **IS NOT** Causal Discovery

Example:

A model might predict:

lung cancer → smoking → heart attack

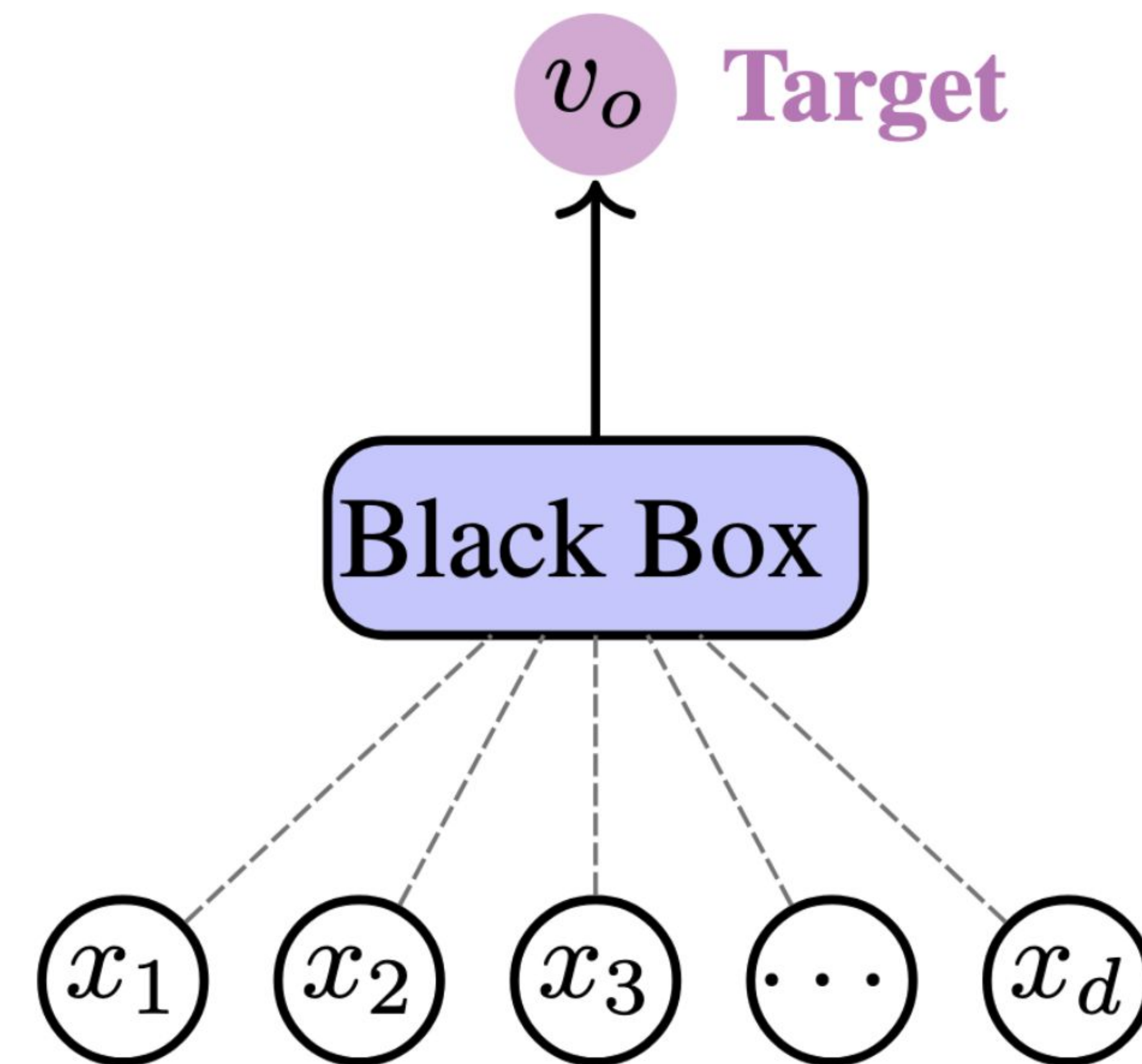
even though in reality:

smoking → lung cancer → heart attack

Problem Statement

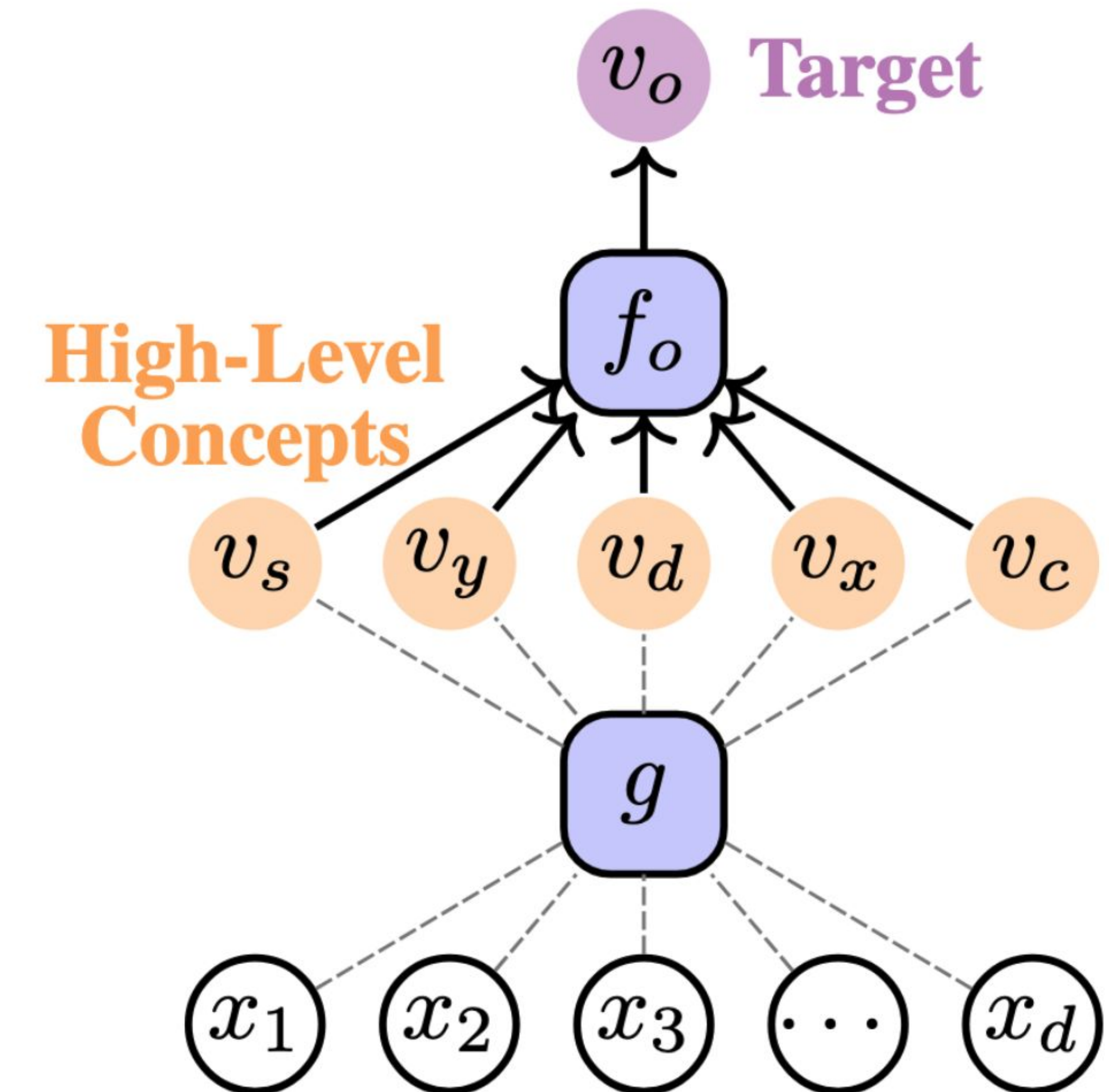
Black Box:

- ✗ Semantically Opaque
- ✗ Causally Opaque



CBMs:

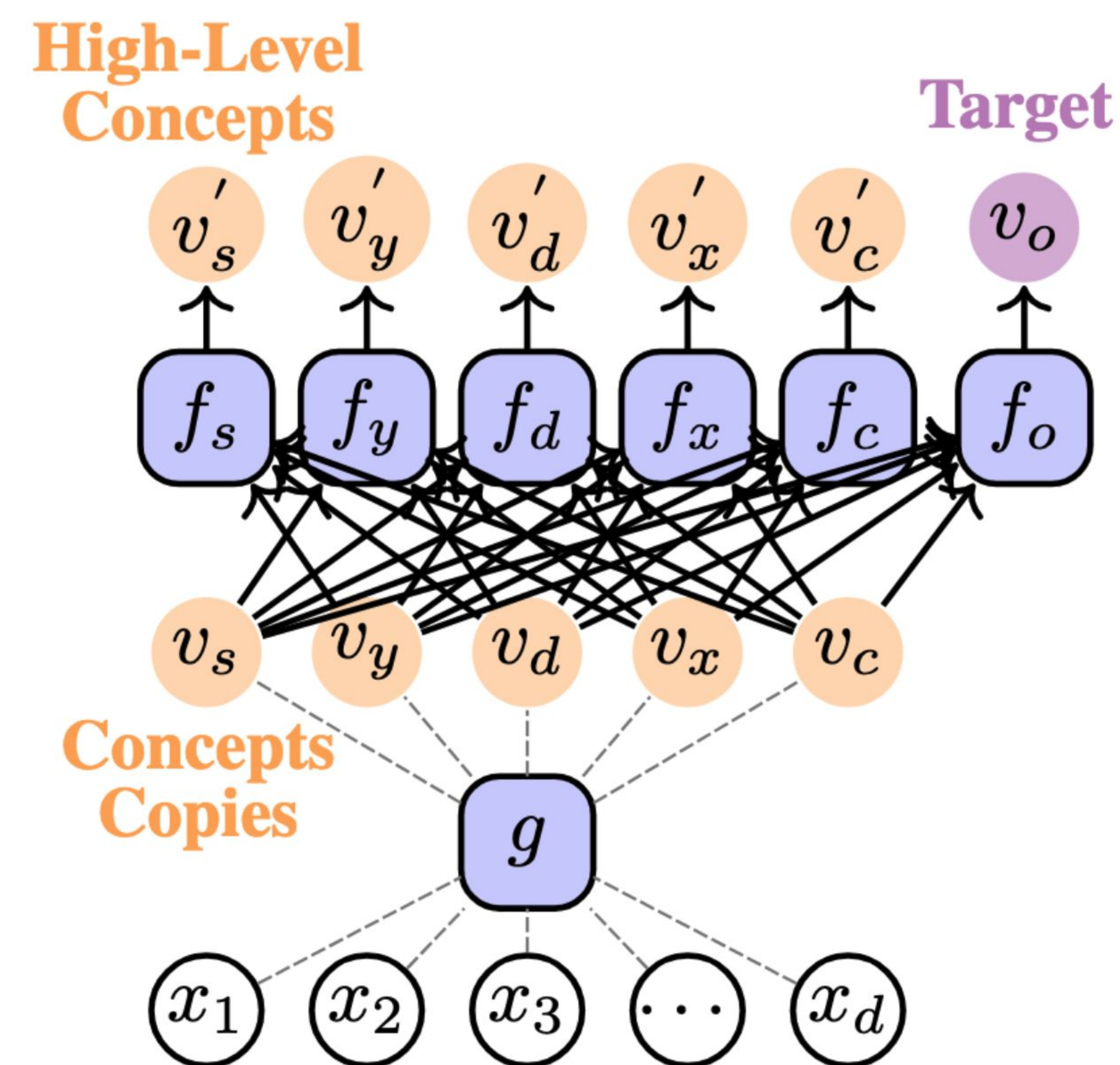
- ✓ Semantically Transparent
- ✗ Causally Opaque



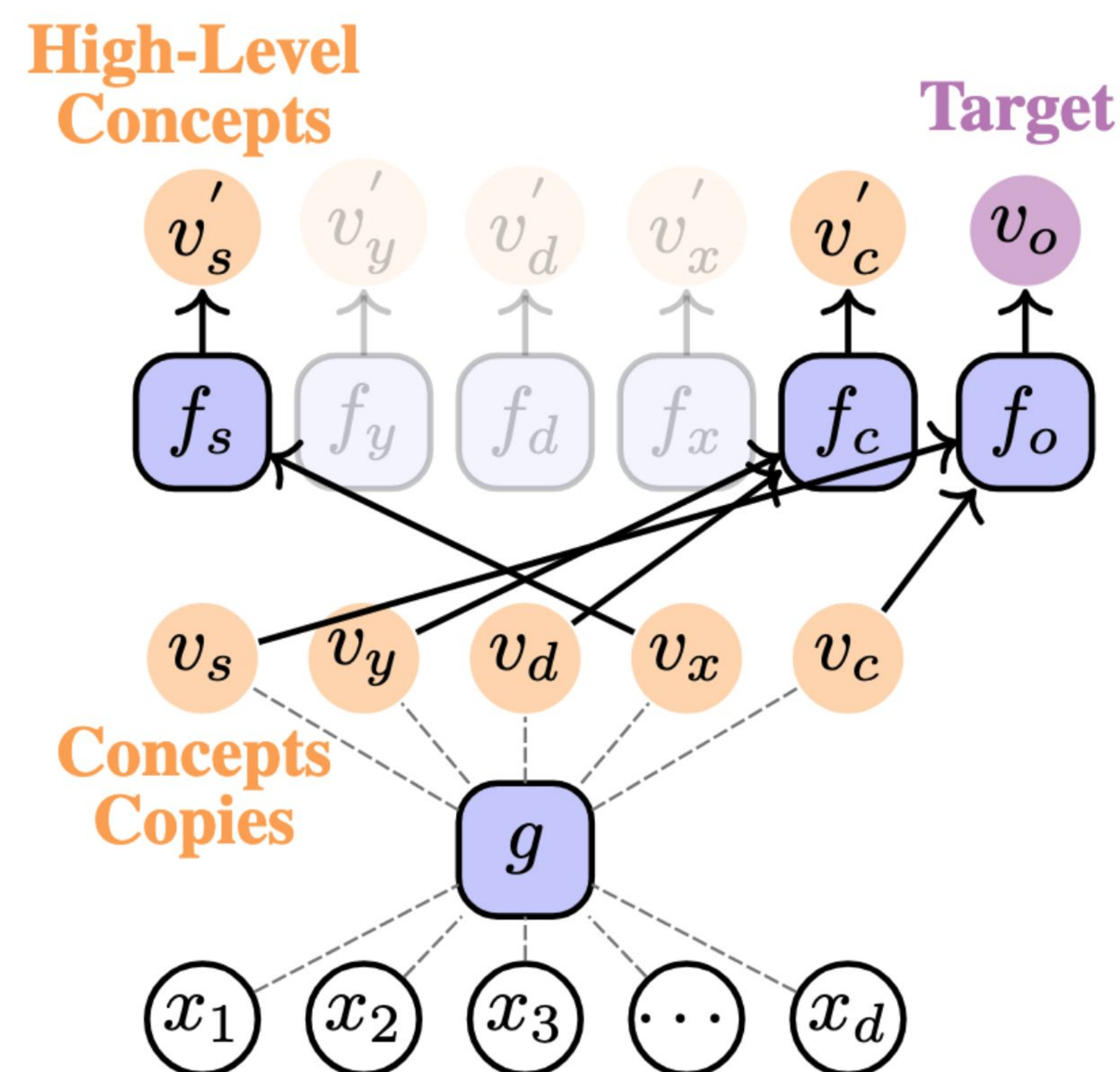
Causal Concept Graph Models

Idea: remove independence assumption from CBMs, add a graph.

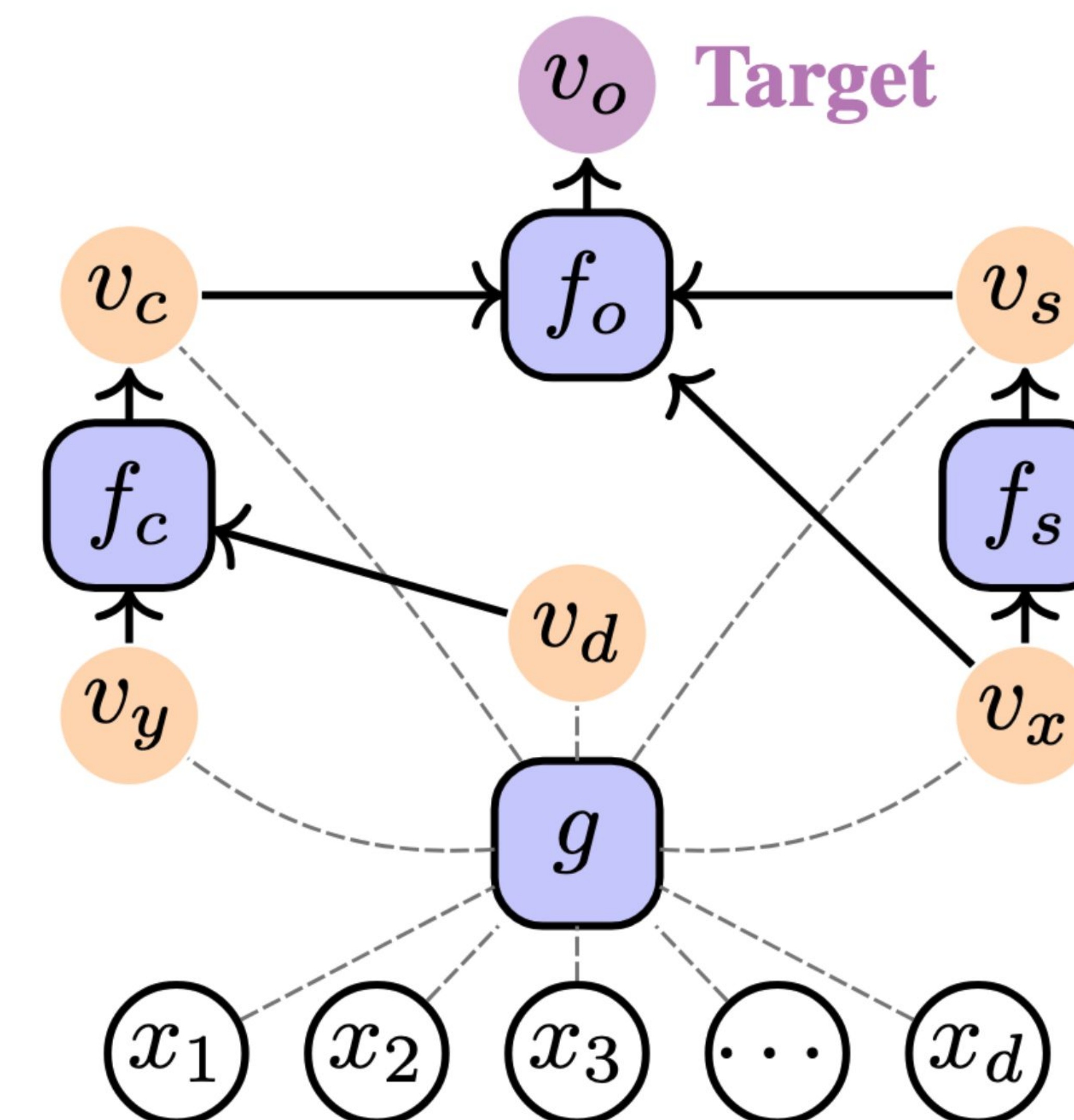
Architecture:
Model all possible dependencies



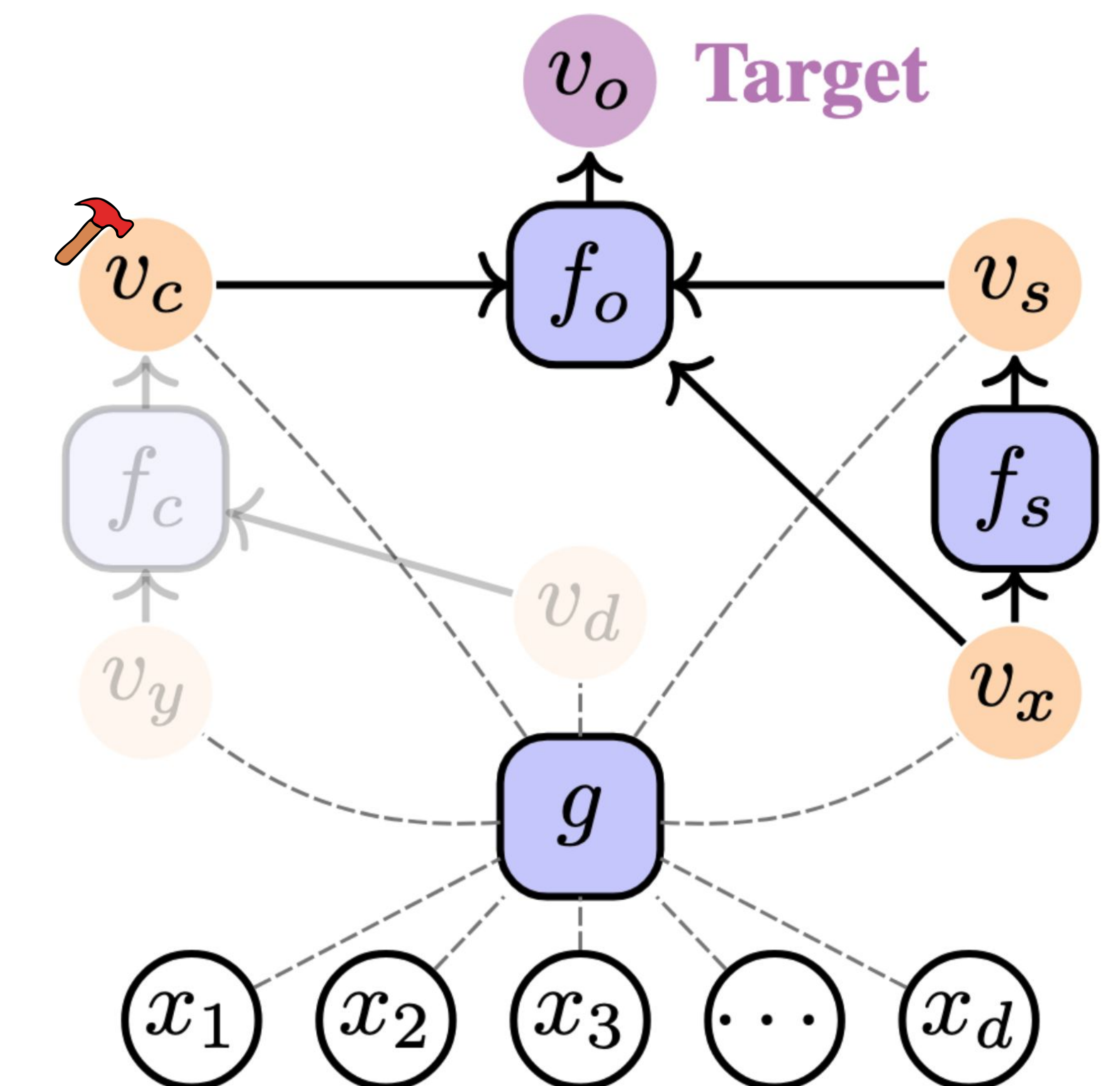
Optimization:
Sparsify connections to a DAG



Inference:
Unroll the connections



Intervention:
Modify the reasoning steps

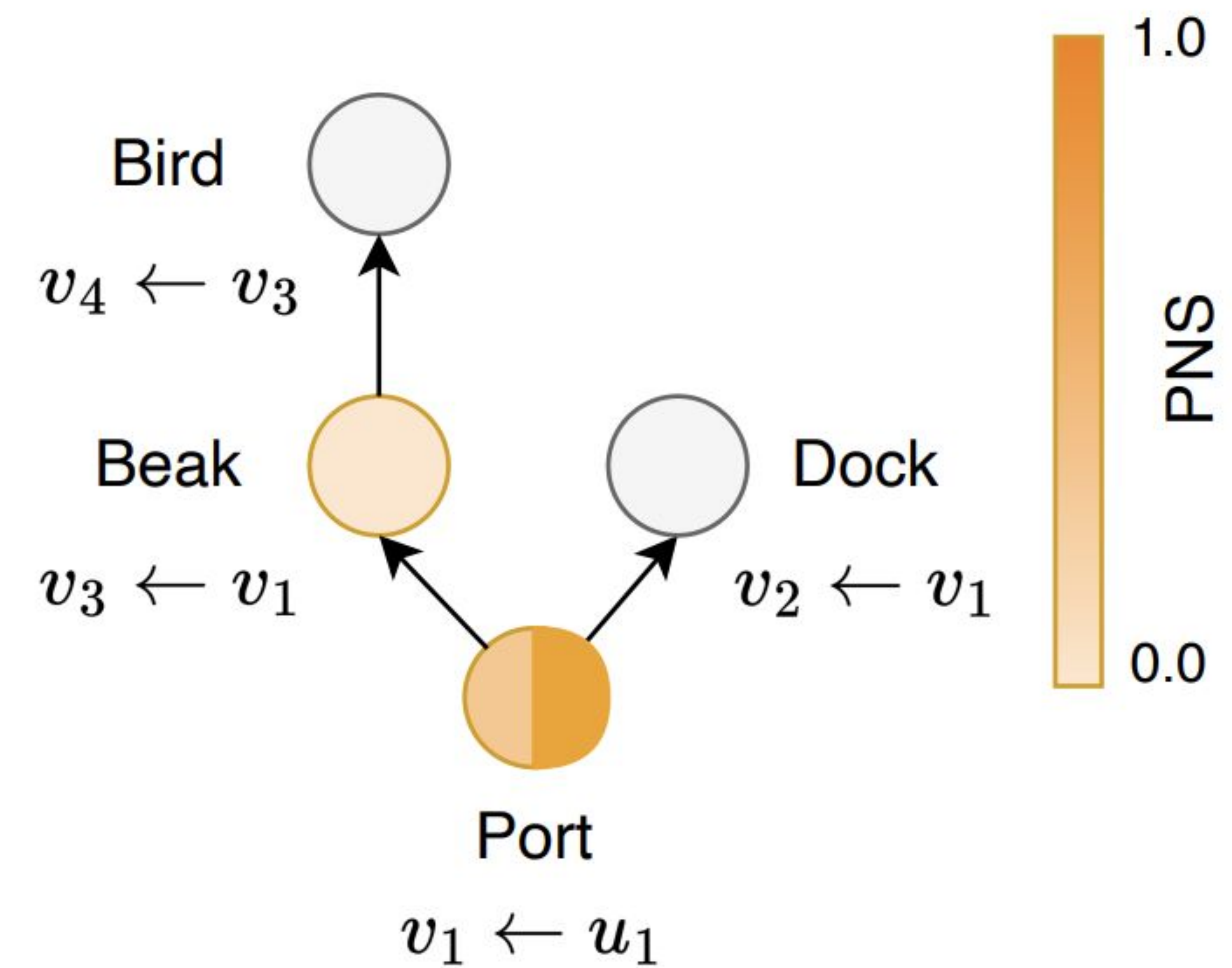
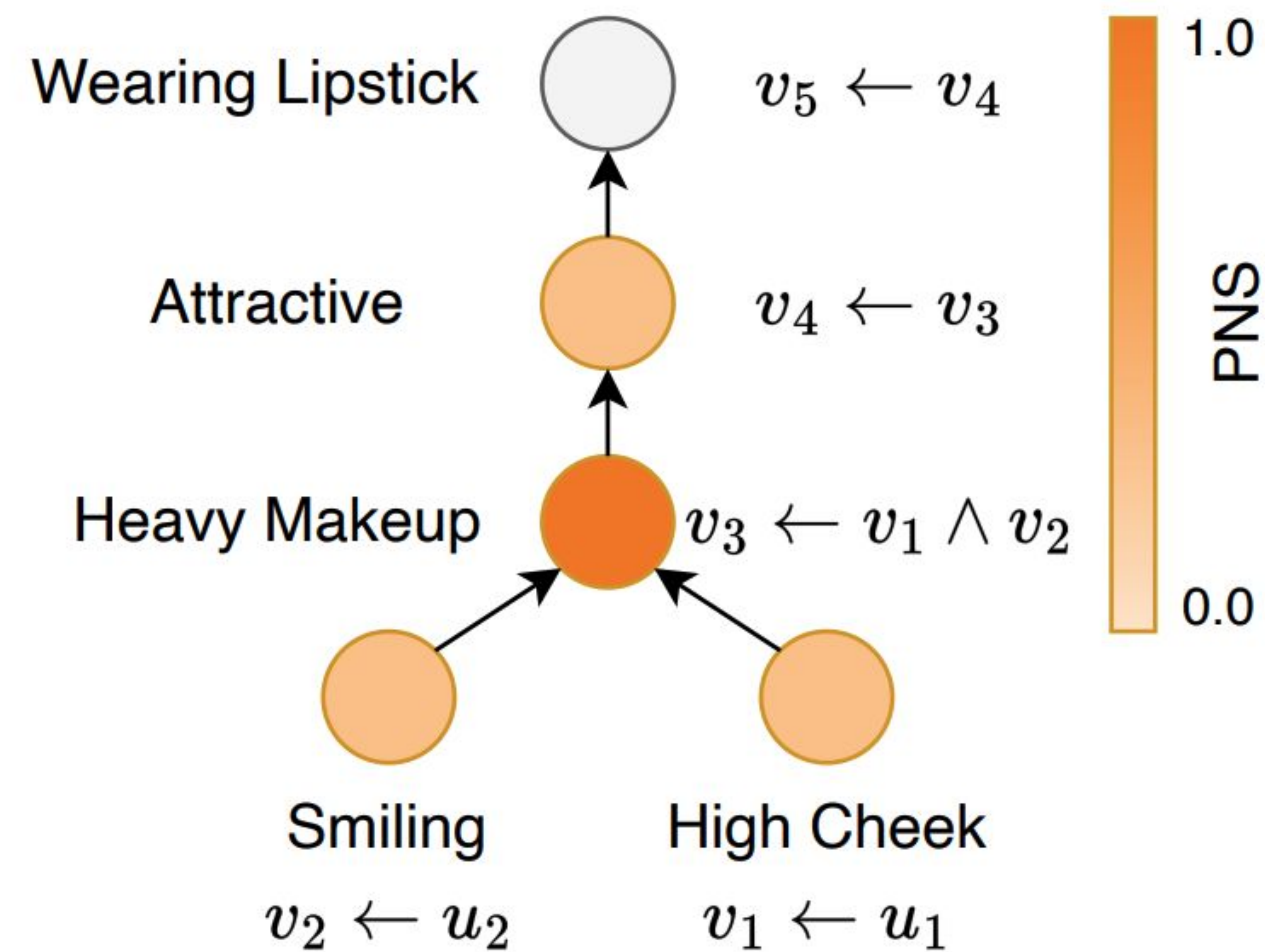


Results - Performance

Label (Concepts and Tasks) Accuracy

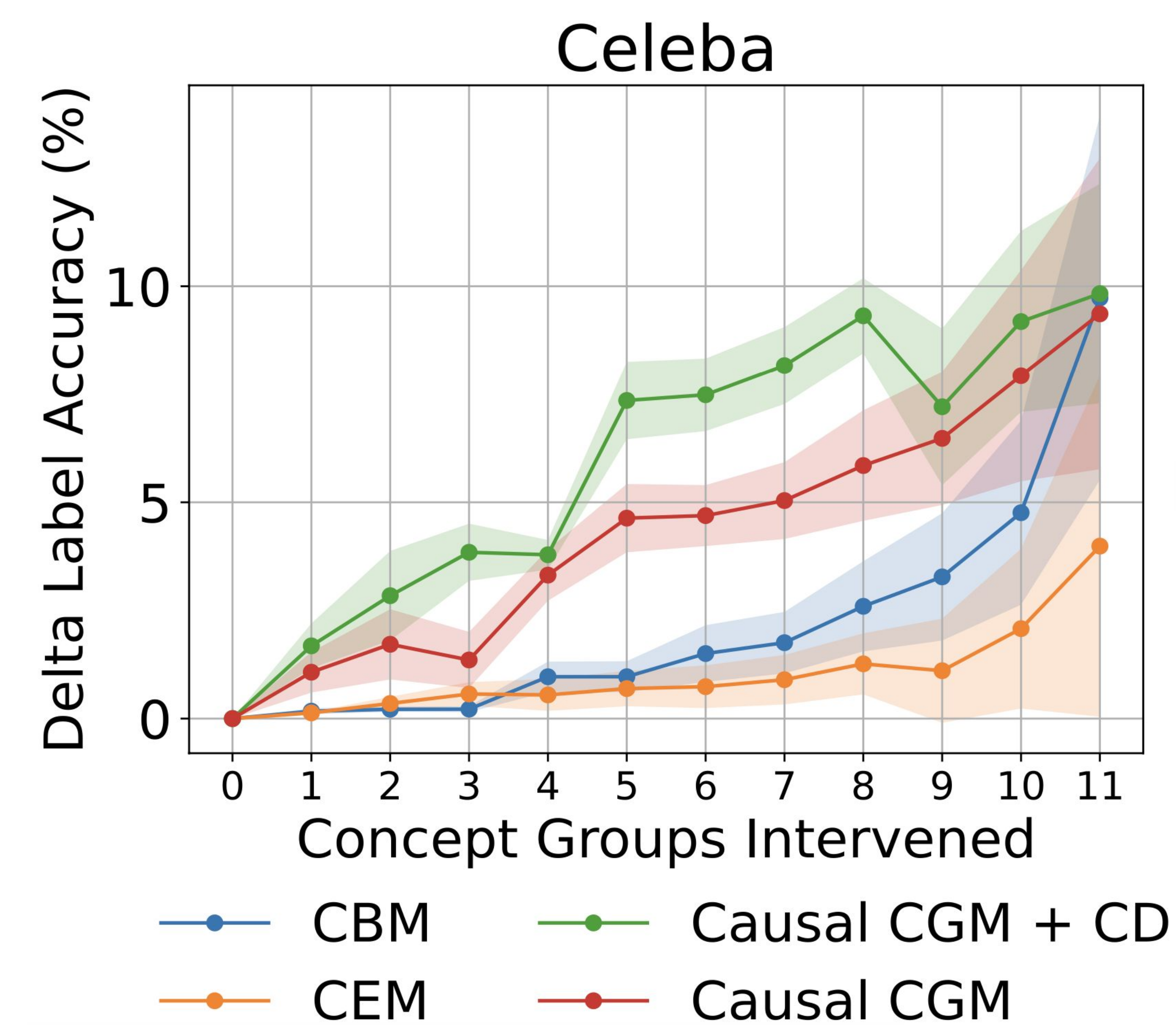
	DSPRITES	CELEBA	CIFAR10
Black box	99.53 $\pm$ 0.05	79.55 $\pm$ 0.14	94.85 $\pm$ 0.03
CBM	99.55 $\pm$ 0.07	79.00 $\pm$ 0.18	92.17 $\pm$ 0.11
CEM	99.48 $\pm$ 0.07	79.17 $\pm$ 0.26	92.04 $\pm$ 0.06
Causal CGM+CD	99.40 $\pm$ 0.15	78.42 $\pm$ 0.42	N/A
Causal CGM	99.44 $\pm$ 0.11	78.23 $\pm$ 0.45	93.32 $\pm$ 0.03

Results - Discovered Graph

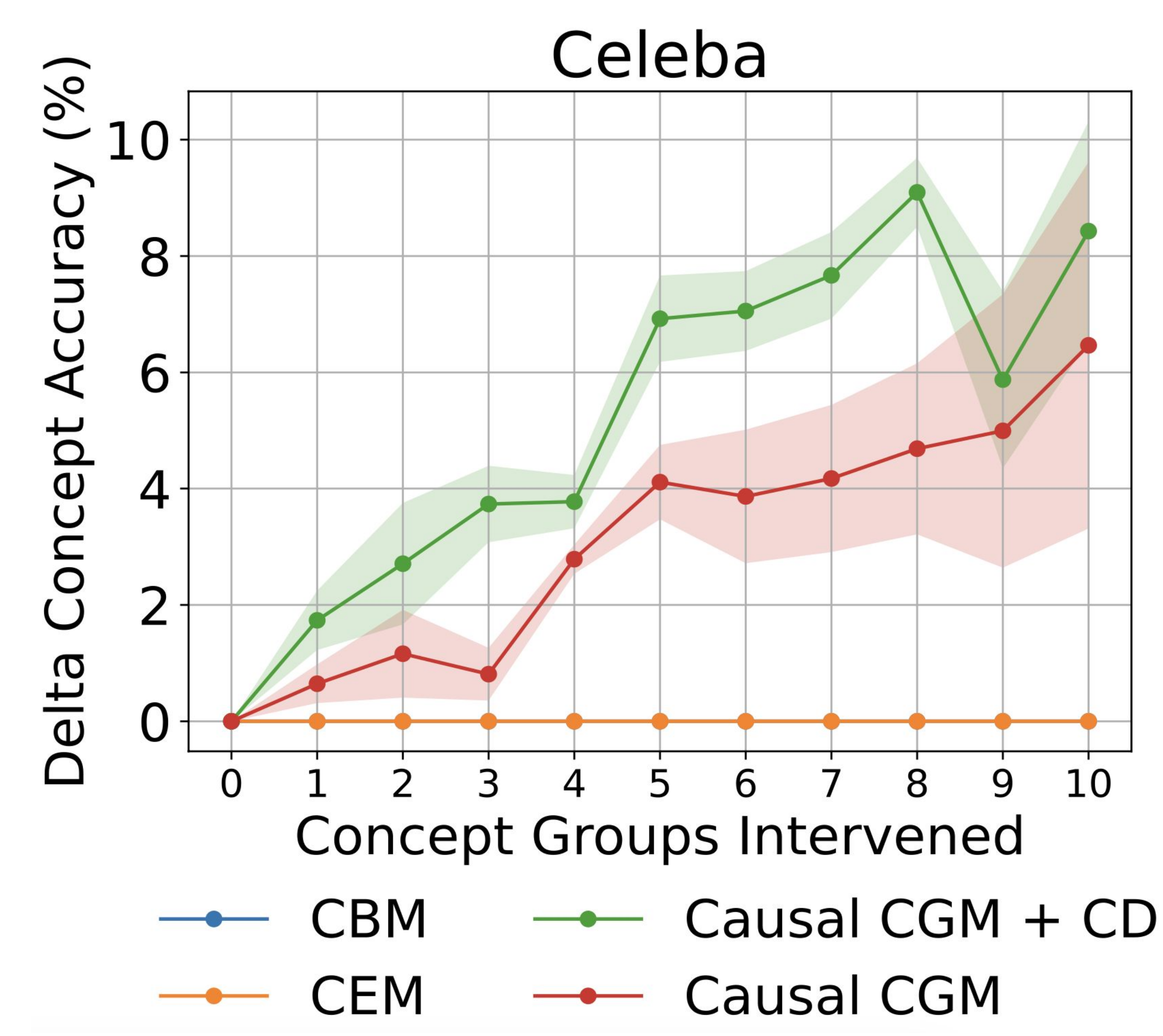


Results - Intervenableability

Label Relative Improvement



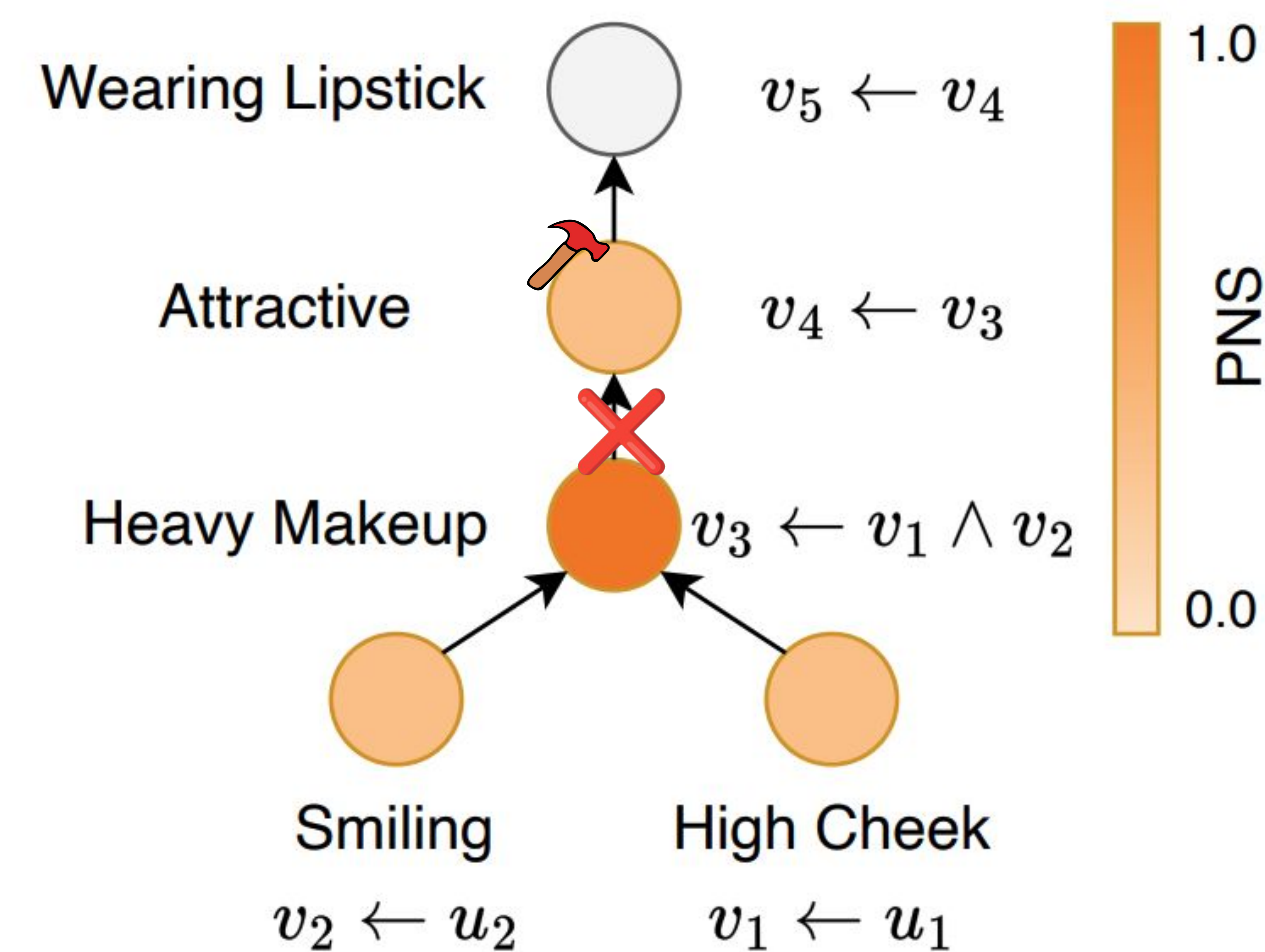
Concept Relative Improvement



Results - Counterfactuals

	Endogenous variables' activations
Initial Predicted State	Smiling, Not Attractive, Mouth Slig, High Cheek, Not Wearing Li, Not Heavy Make, Not Wavy Hair, Not Big Lips, Not Oval Face, Not Makeup, Not Fem Model
What if I would wear lipstick?	Smiling, <u>Attractive</u> , Mouth Slig, High Cheek, Wearing Li , <u>Heavy Make</u> , Not Wavy Hair, Not Big Lips, Not Oval Face, Not Makeup, Not Fem Model
What if I would wear lipstick and also makeup?	Smiling, <u>Attractive</u> , Mouth Slig, High Cheek, Wearing Li , <u>Heavy Make</u> , Not Wavy Hair, <u>Big Lips</u> , Not Oval Face, Makeup , <u>Fem Model</u>

Results - Make two variables causally independent



Residual Concept Causal Effect

	DSPRITES	CELEBA	CIFAR10
Black box	N/A	N/A	N/A
CBM	100.00±0.70	97.84±2.13	100.00±0.00
CEM	100.00±4.62	106.00±0.50	100.00±0.00
Causal CGM+CD	0.00±0.00	0.00±0.00	N/A
Causal CGM	0.00±0.00	0.00±0.00	0.00±0.00



CAUSAL CONCEPT GRAPH MODELS: BEYOND CAUSAL OPACITY IN DEEP LEARNING

Gabriele Dominici, Pietro Barbiero, Mateo Espinosa Zarlenga, Alberto Termine, Martin Gjoreski, Giuseppe Marra, Marc Langheinrich*

