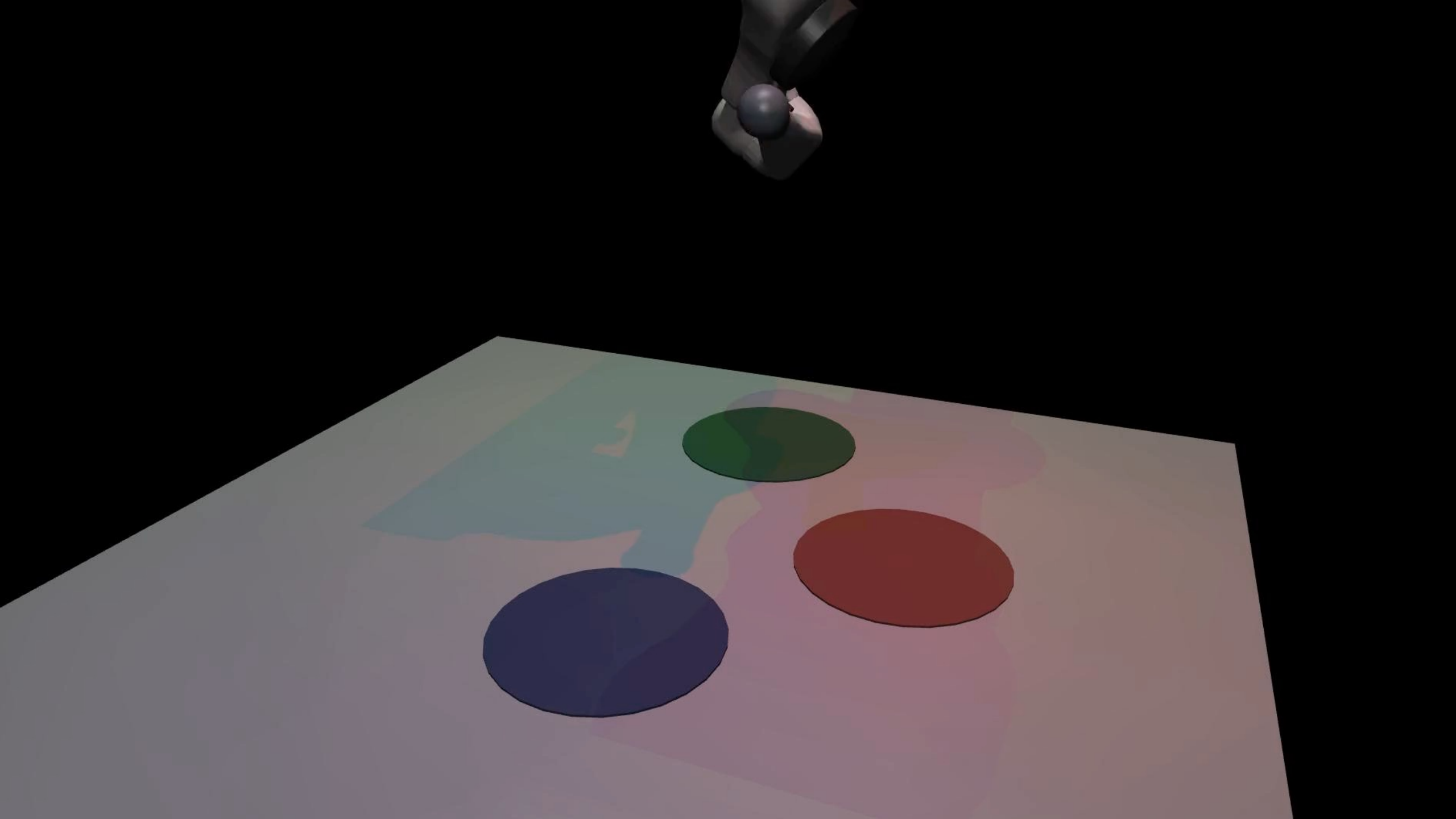
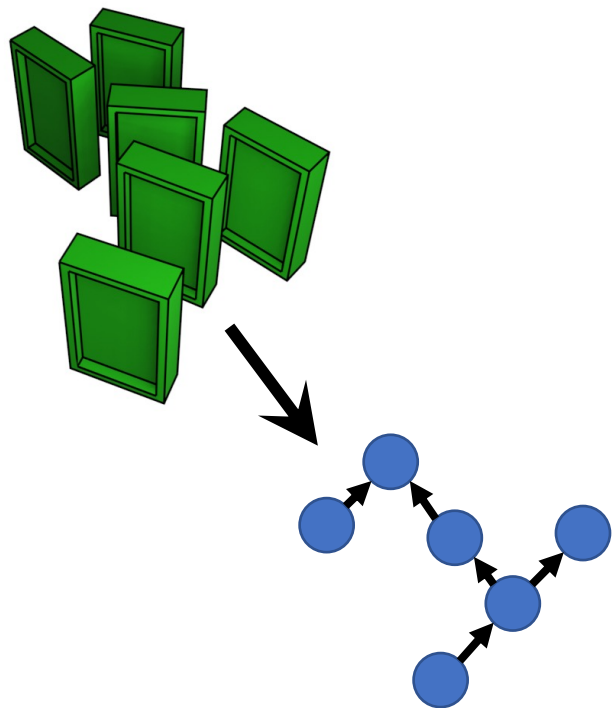


# Weakly supervised causal representation learning

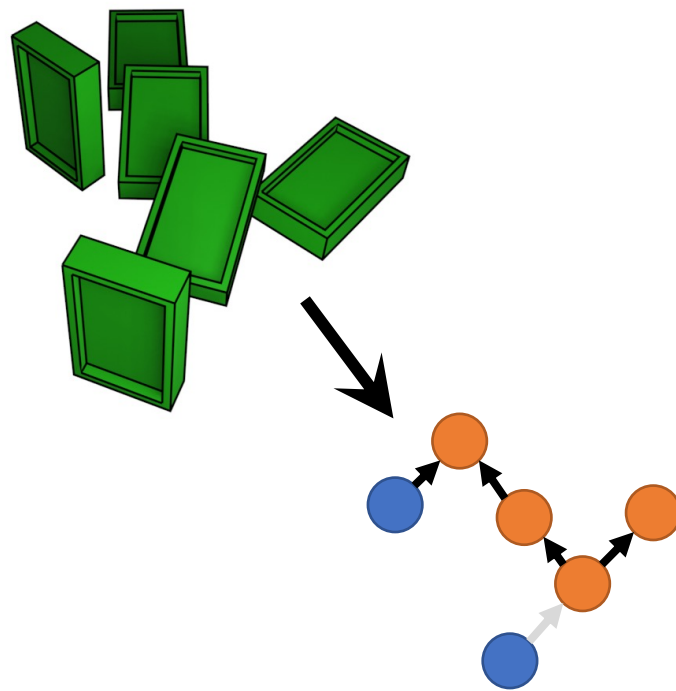
**Johann Brehmer**

Qualcomm Technologies Netherlands B. V.

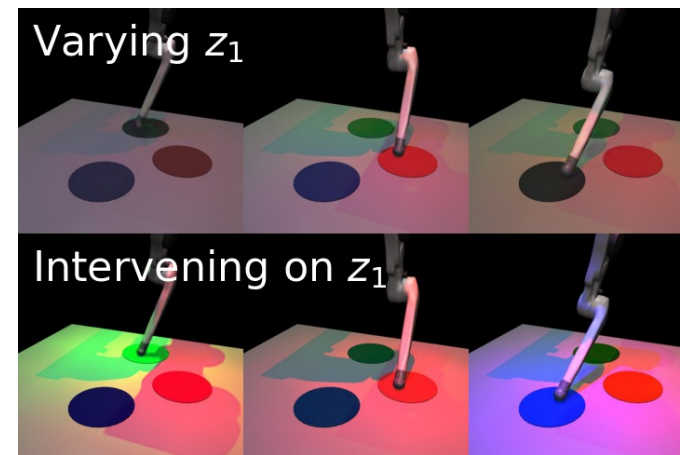




Can we **learn causal variables & causal structure from pixels**, without labels?



We prove: this is possible with **weak supervision**, when observing effects of interventions

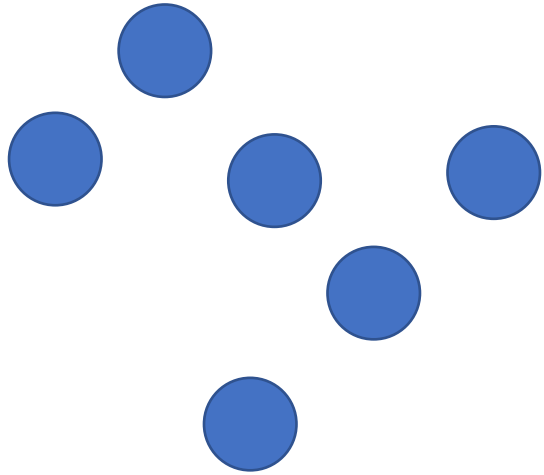


In practice, **implicit latent causal models** can identify the causal structure in image datasets

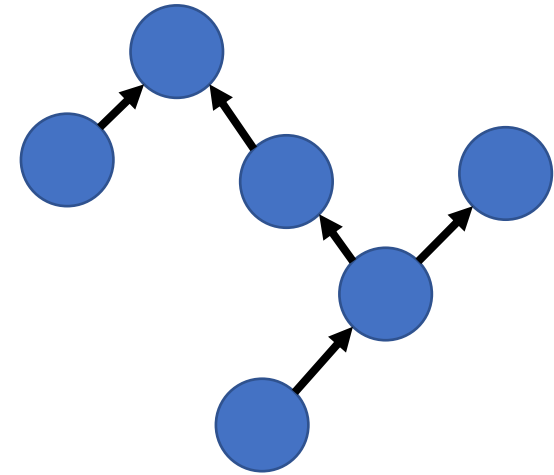
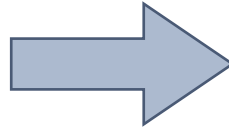
Problem

**Can we learn causal representations  
from pixels?**

# Causal discovery / inference



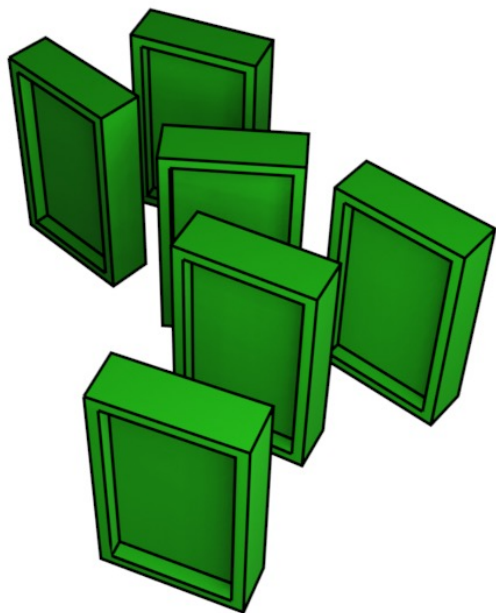
Given: dataset in terms of  
**high-level causal variables**



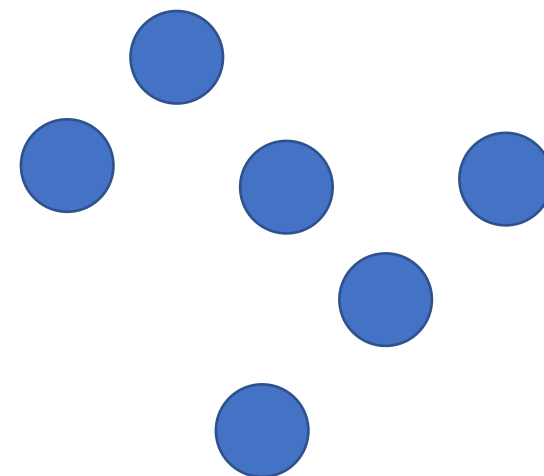
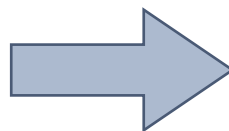
Goal: learn the  
**causal structure**

**But: what if we don't observe the causal variables?**

# Disentangled representation learning



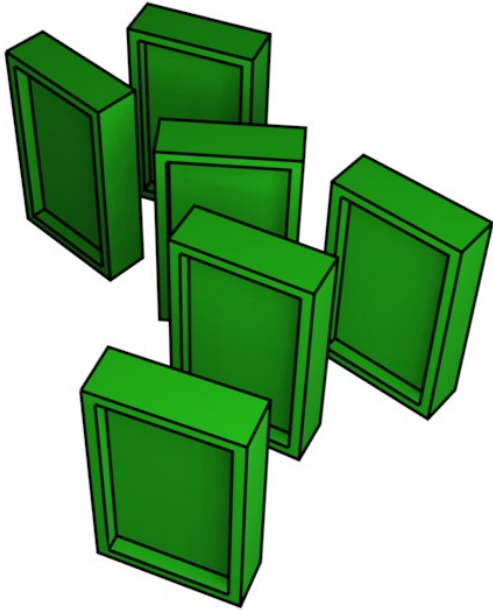
Given: **low-level, unstructured data representation**  
(e.g. pixels)



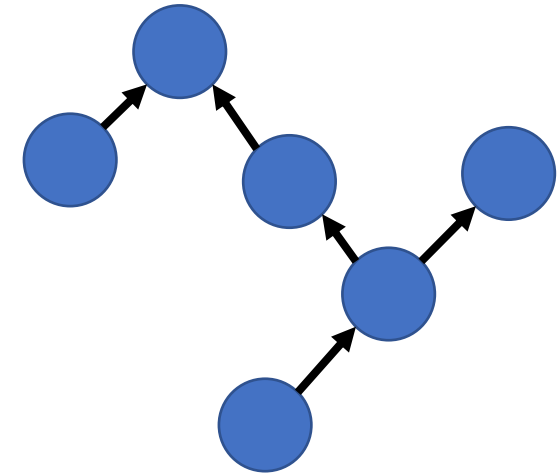
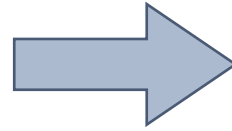
Goal: learn encoder to  
**high-level variables**  
(e.g. object positions, states, ...),  
usually **assuming independence**

**But: useful high-level concepts are rarely independent**

# Causal representation learning

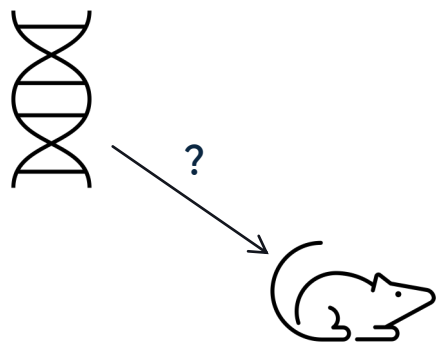


Given: **low-level, unstructured data representation**  
(e.g. pixels)

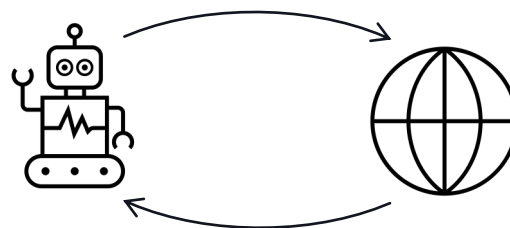


Goal: learn encoder to  
**high-level variables**  
(e.g. object positions, states, ...)  
**and their relations /**  
**causal structure**

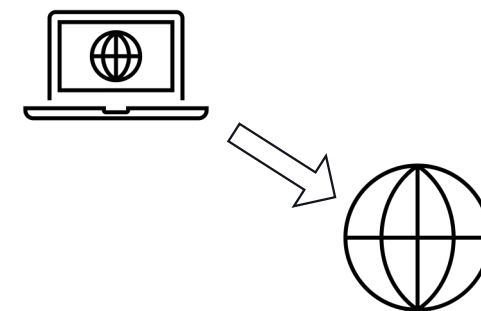
# Why learn causal representations?



Causal structure may be of **scientific interest**



Causal representations are **abstractions** that may be **useful for planning**



Causal models may be more **robust to changes**

Arguably, these potential benefits have not yet been clearly demonstrated

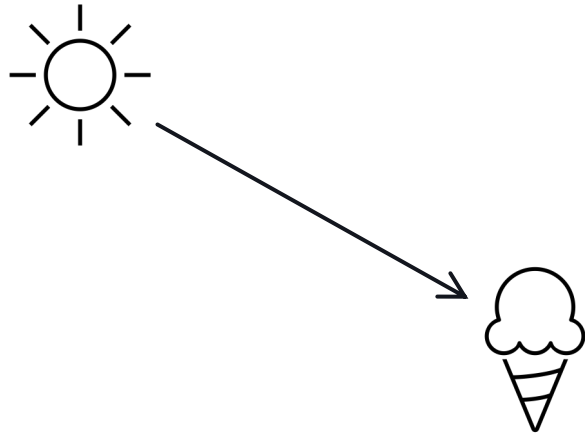
[Recent review: B. Schölkopf et al, "Towards causal representation learning", IEEE Advances in Machine Learning and Deep Neural Networks 2021]



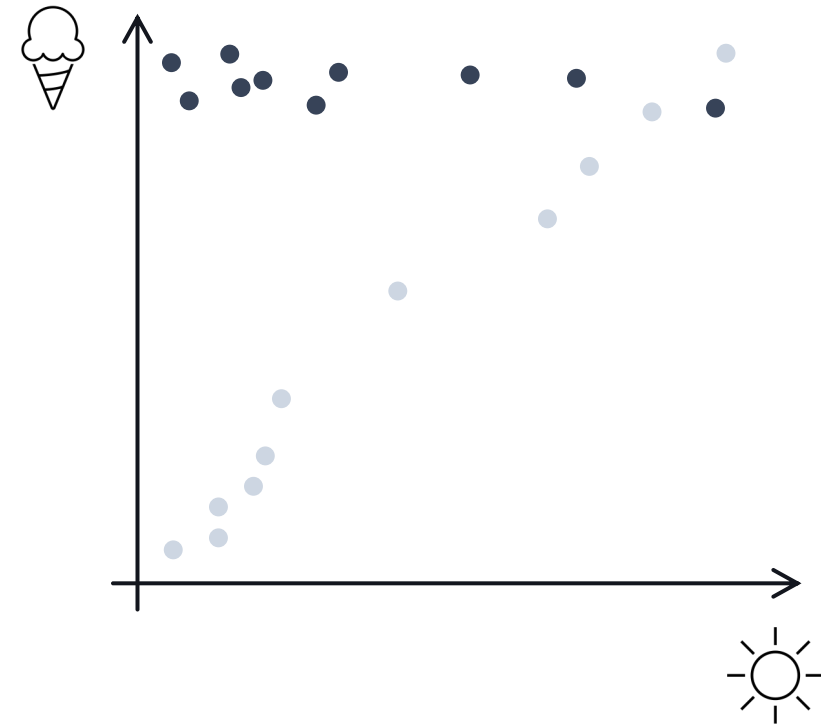
Background

**Causality and identifiability**

# Causality



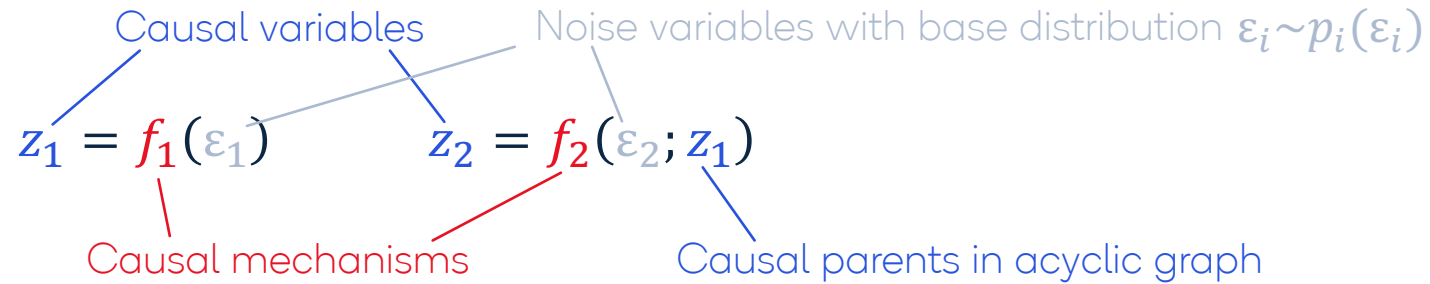
Semantically, causal models label relations between random variables as **cause-effect relations**



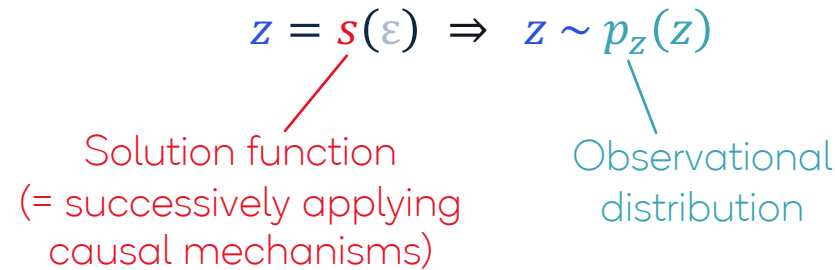
Functionally, causal models describe **probability distributions and how they change** under changing conditions

# Structural causal models (SCMs)

- SCM:



- Solution:



- Interventions:



# Identifiability

- An representation / SCM  $\mathcal{M}$  is **identifiable** if

$$p_{\mathcal{M},x}(x) = p_{\mathcal{M}',x}(x) \Rightarrow \mathcal{M} \sim \mathcal{M}'$$

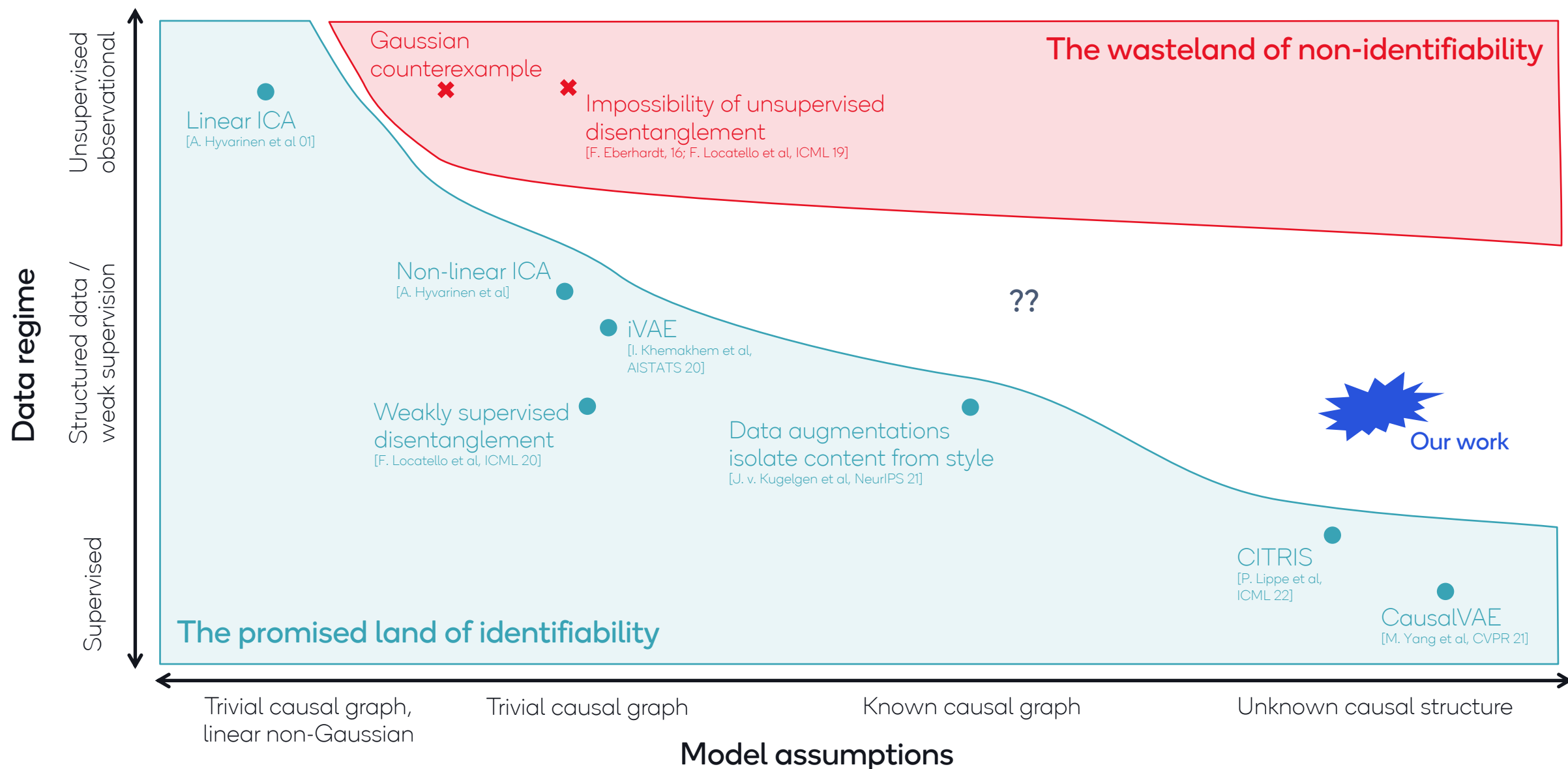
Any two model  
(from some family)

Data regime  
(e.g. observational  
distribution on pixel level)

Equivalence relation  
(e.g. same up to  
permutations)

- Identifiability means we can **find ground-truth causal structure** through maximum-likelihood training
  - if it is within the specified model family
  - up to the equivalence relation
  - in the limit of infinite data
  - assuming perfect training

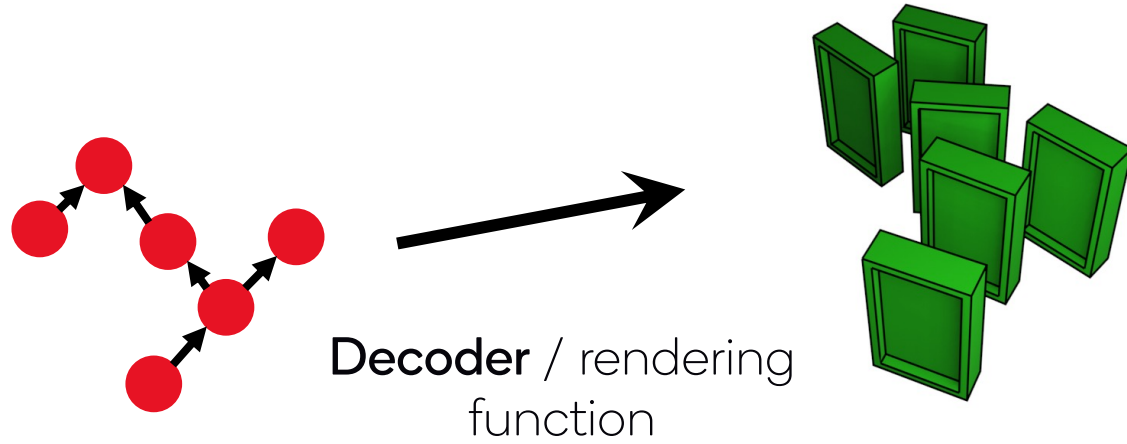
# When are causal representations are identifiable?



Theory

**Causal representations can be identified from weak supervision**

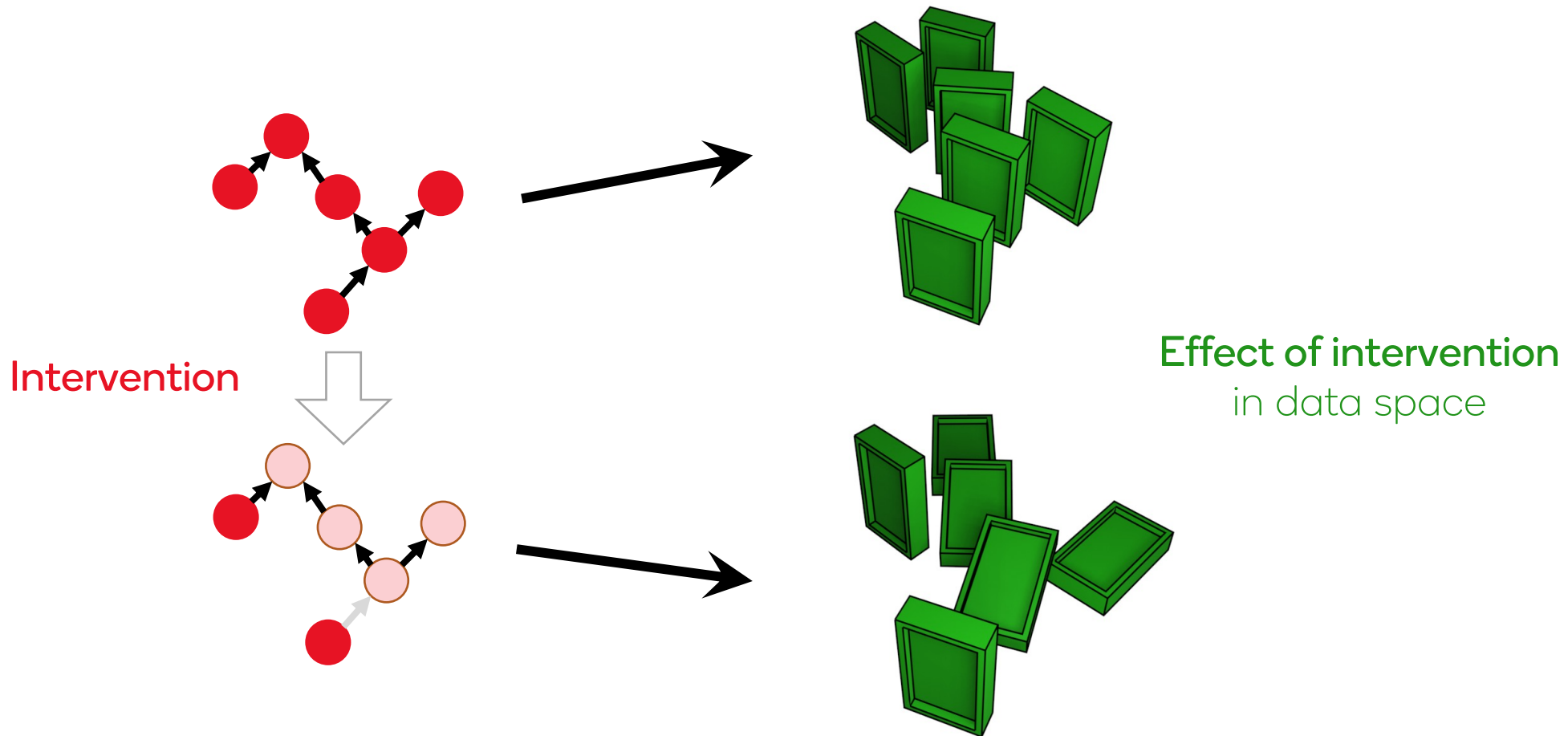
# Latent causal model



High-level variables with  
a structural causal model  
between them

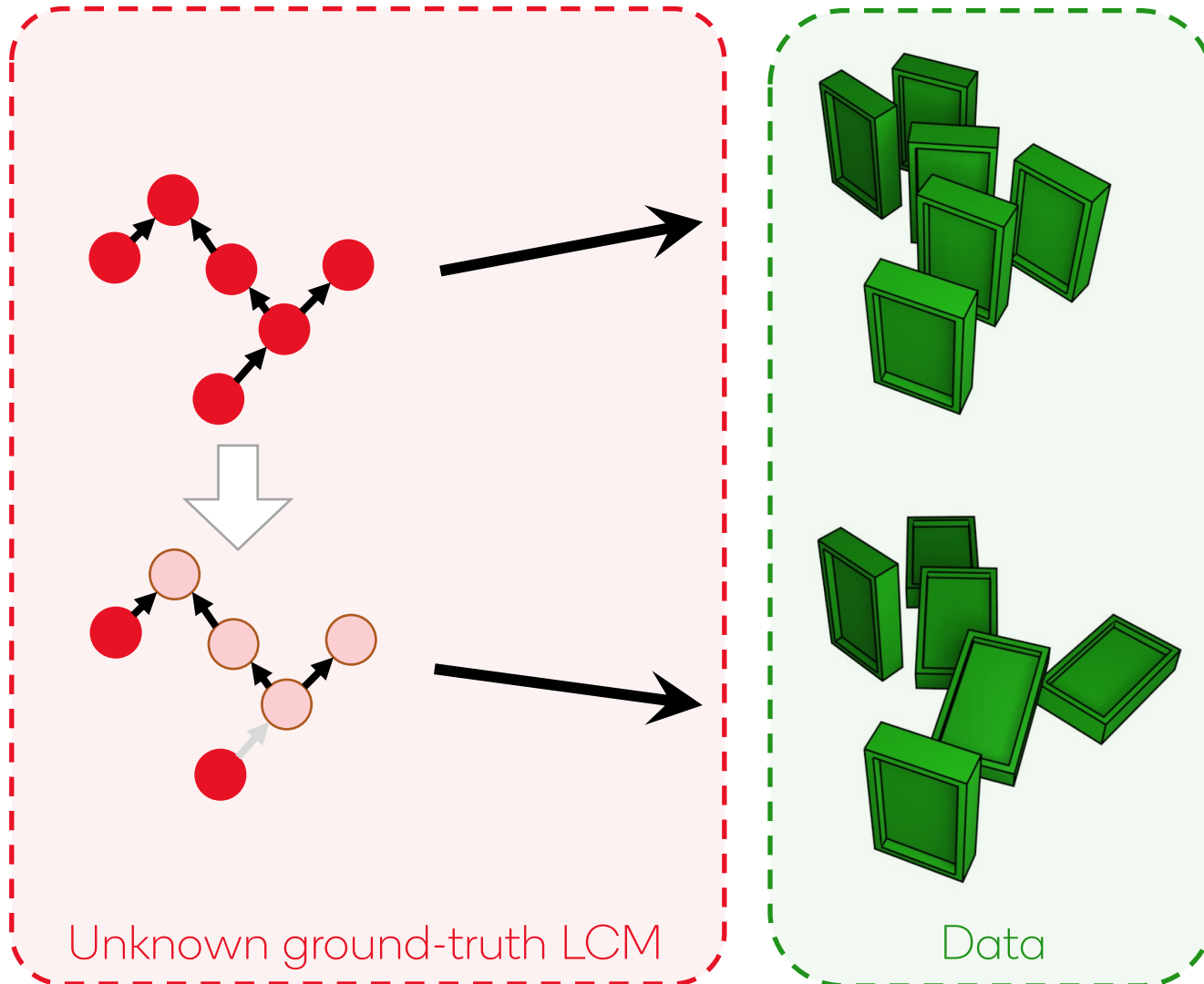
Low-level data (pixels)

# Interventions



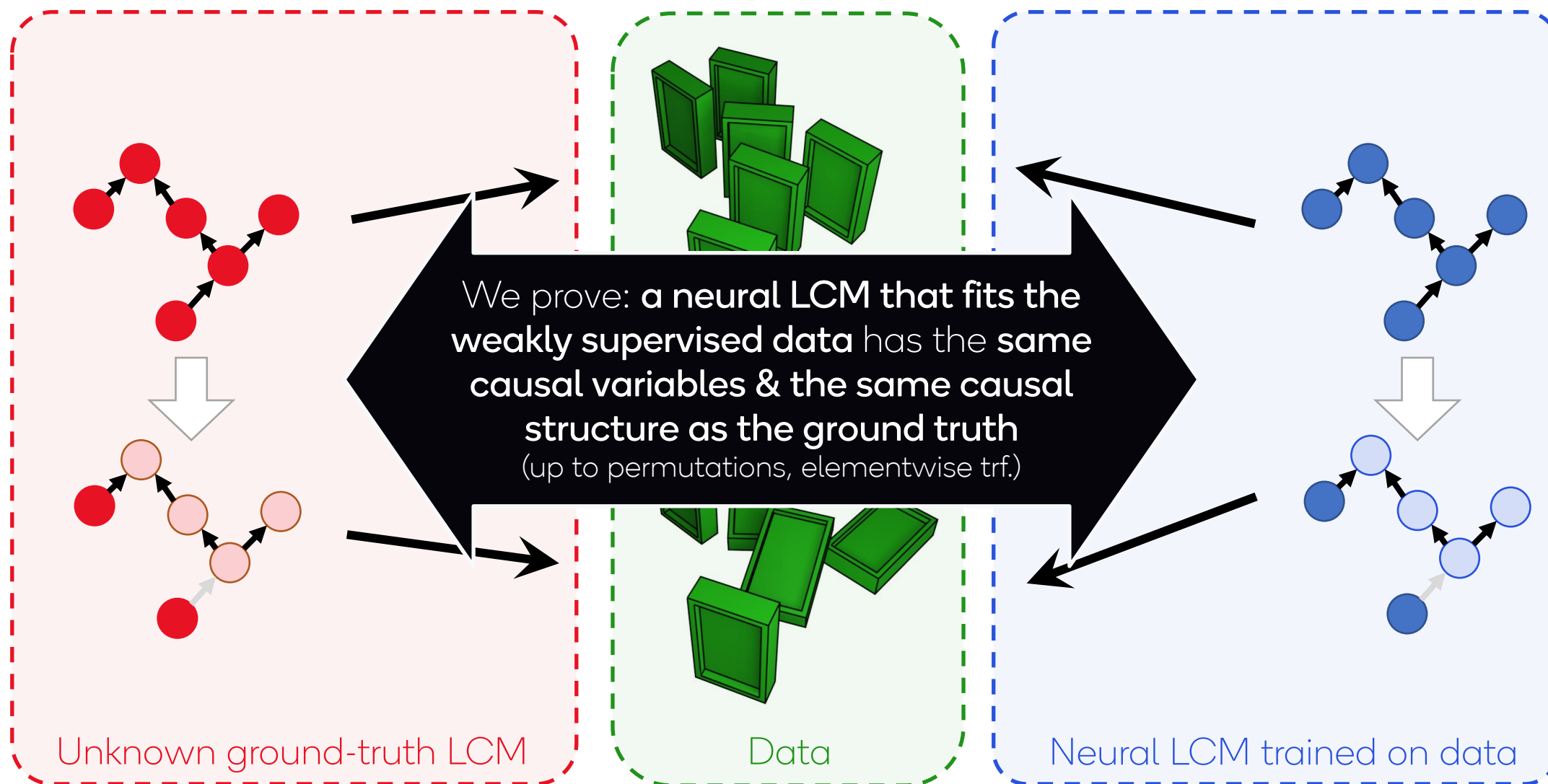


# Weakly supervised data setting



- We assume access to **data pairs of the system before and after interventions**
  - Equivalent to counterfactuals
  - Causal abstraction of time-series data
- Otherwise, **no labels**
  - Only pixel-level data is observed
  - Intervention targets are unknown

# Identifiability theorem

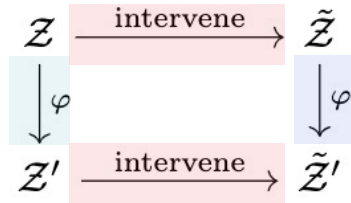


# Proof sketch

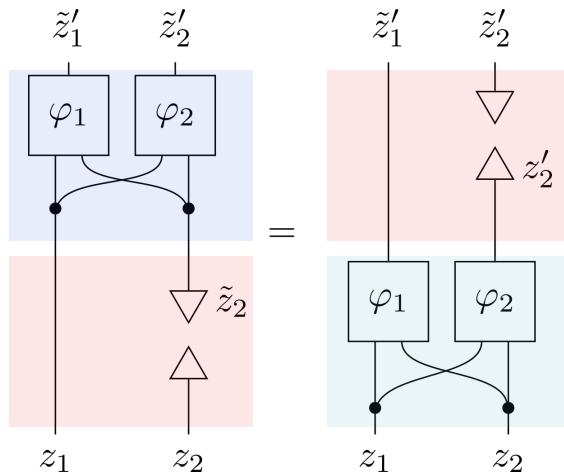
1. Consider two LCMs with causal variables  $\mathbf{z}$  and  $\mathbf{z}'$ , both matching the data.

Define  $\varphi : \mathbf{z} \rightarrow \mathbf{z}'$ .

2. Interventions commute with  $\varphi$ :



3. We assume perfect interventions. Then then  $\tilde{z}'_i$  is independent of  $\mathbf{z}_j$ . For 2 variables:



4. We assume  $\mathbb{R}$ -valued variables. Statistical independence then implies functional independence. Thus,  $\varphi_i(\mathbf{z}_i, \mathbf{z}_j)$  must be constant in  $\mathbf{z}_j$ .
5. Since this holds for any  $i$ ,  $\varphi$  must be a permutation plus elementwise transformations.
6. Finally, we can show that the causal graphs and intervention targets in the two models are consistent with this transformation.
7. Thus the two models are isomorphic.

# Assumptions

Assumption

**Weakly supervised data is available**

**Causal variables are  $\mathbb{R}$ -valued**

**Causal mechanisms are diffeomorphic**

**No hidden confounders**

**Decoder is deterministic**

**Interventions are perfect**

(Post-intervention values of intervention targets are independent of pre-intervention state)

**Interventions are complete**

(The dataset contains interventions on any single causal variable)

Possible relaxation

Maybe (work in progress)

Maybe (work in progress)

Difficult

Difficult

Plausible (as in iVAE)

Difficult (counterexamples)

Relaxation to n-target interventions plausible  
(incomplete interventions → partial identifiability)

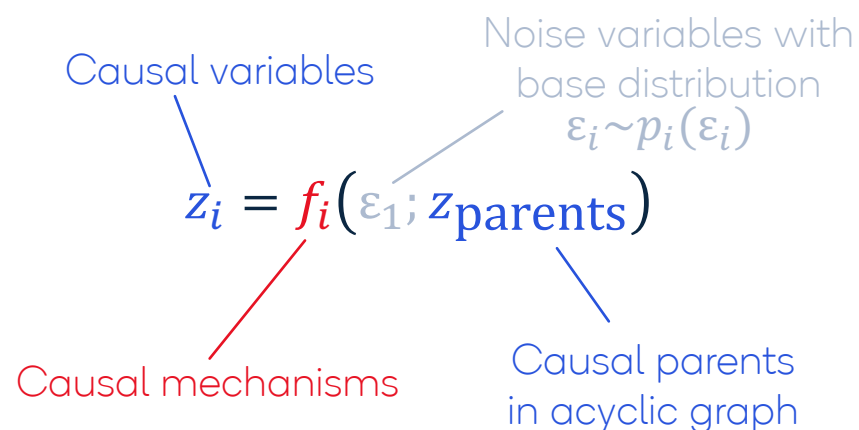
Practice

**Implicit is better than explicit**

# Explicit and implicit representations of causal structure

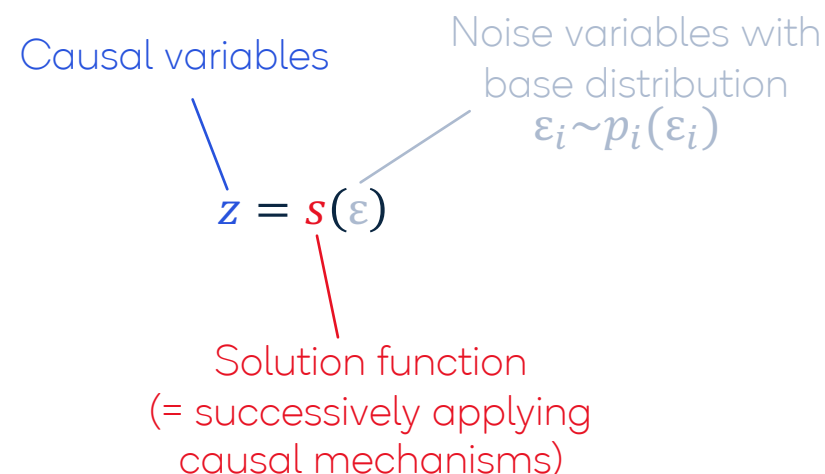
## Explicit representation

through graph & causal mechanisms:



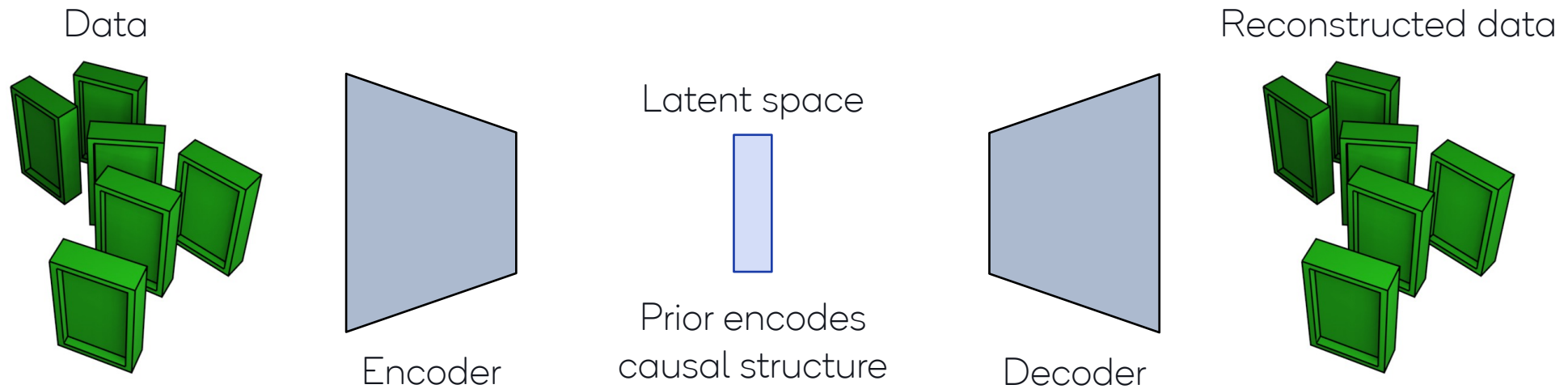
## Implicit representation

through solution function:

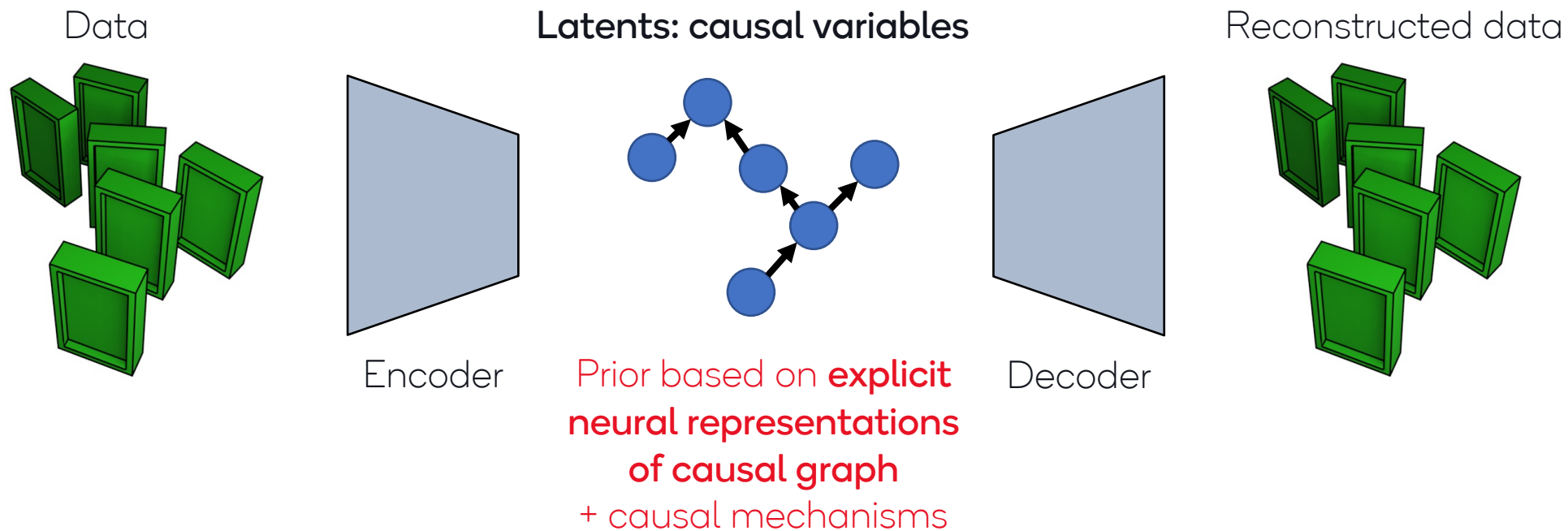


Under our assumptions, explicit and implicit representation **contain the same information**

# Operationalizing latent causal models

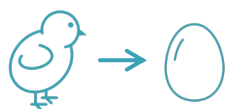


# Explicit latent causal models





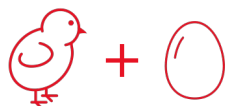
# Explicit latent causal models in practice



Easy to learn graph given representations



Easy to learn representations given graph

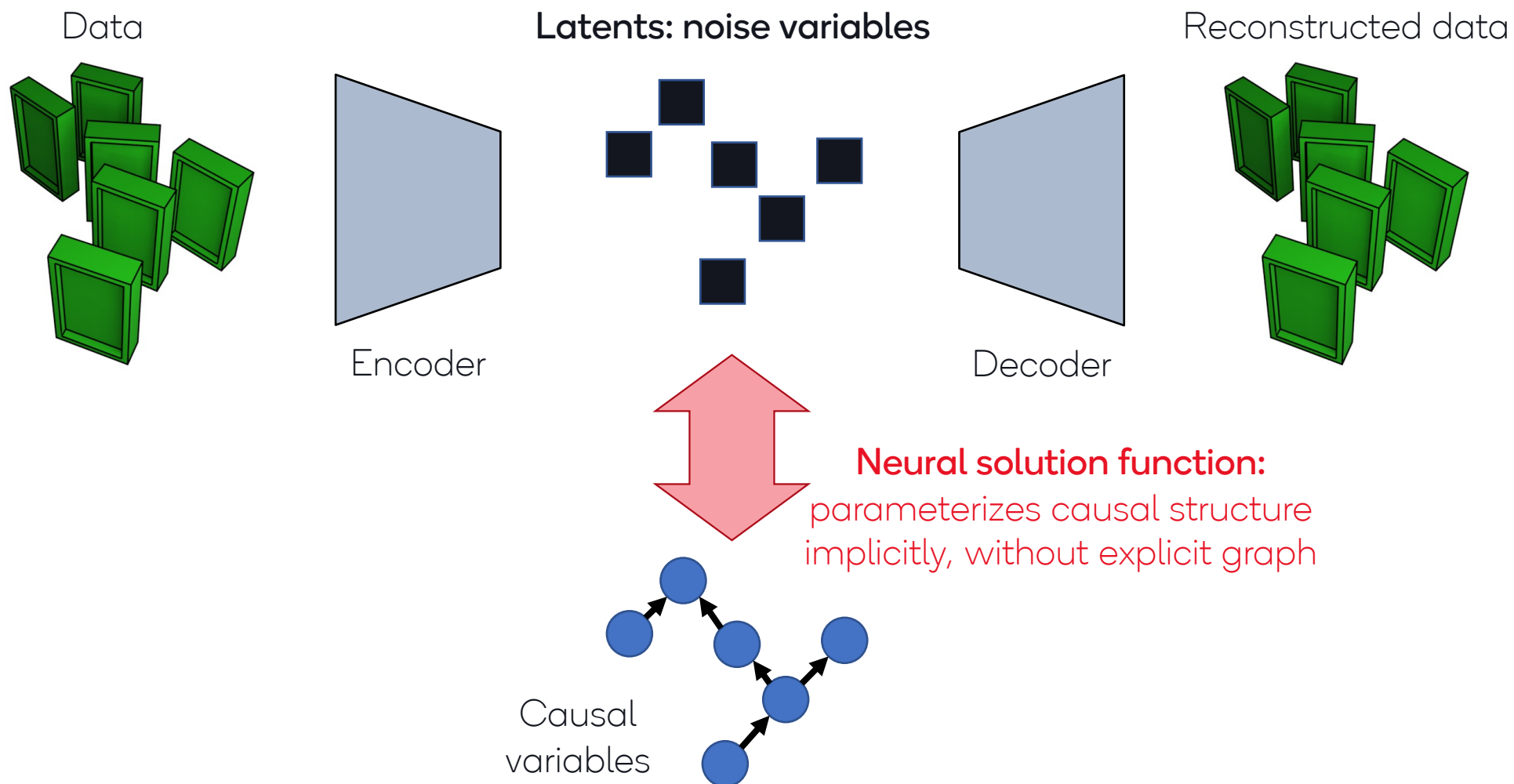


**Difficult to learn graph and representation simultaneously**

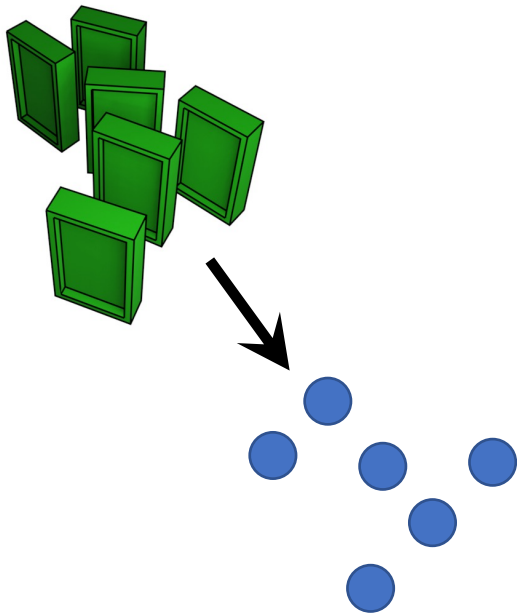
(Evidence for **local minima** in the loss landscape corresponding to wrongly oriented graph edges)

⇒ don't learn explicit graphs if you don't have to

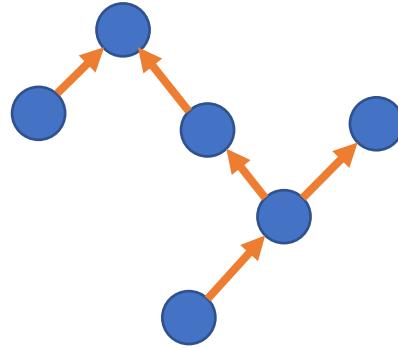
# Implicit latent causal models



# What can you do with ILCMs?

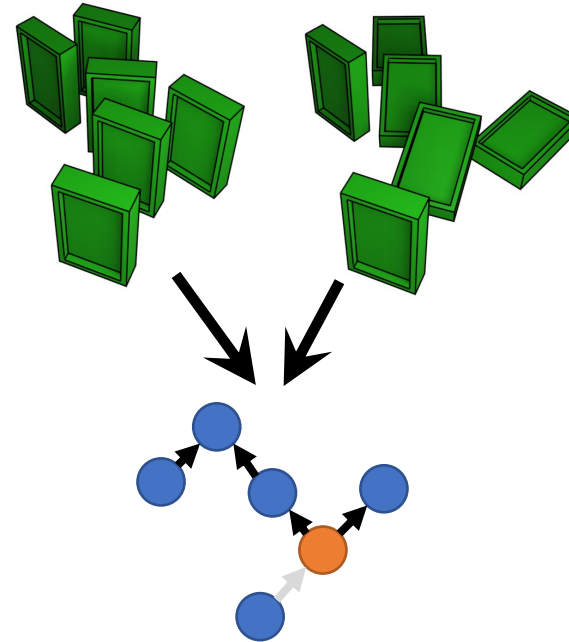


**Map pixels to causal variables**

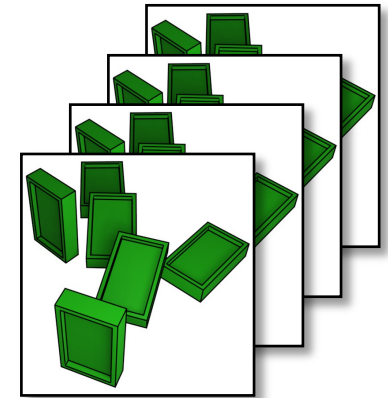


**Find the causal graph**

- ILCM-E: with off-the-shelf causal discovery algorithm ENCO
- ILCM-H: with our new heuristic



**Infer interventions from data pairs**



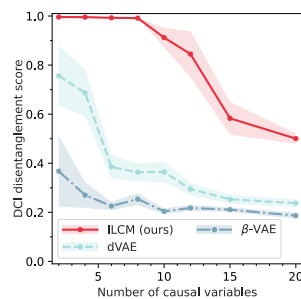
**Generate observational, interventional, and counterfactual data**

Experiments

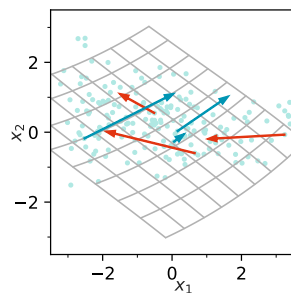
**Things work, mostly**

# Experiments

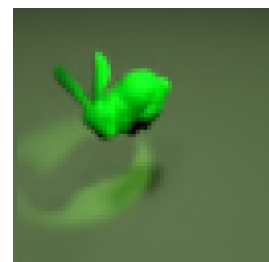
Complexity of causal system



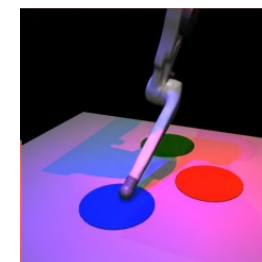
Scaling experiment



2D toy example



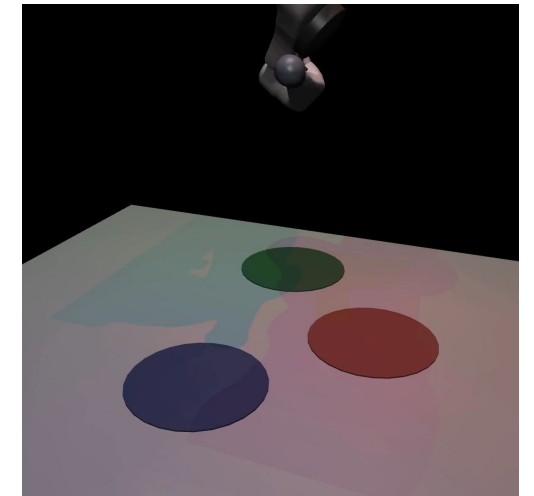
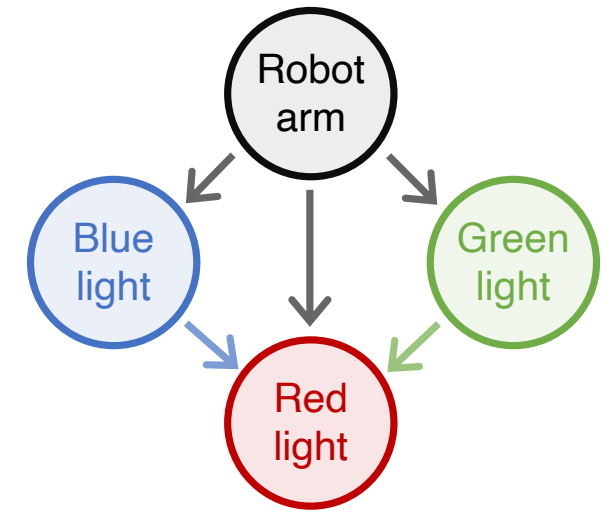
Causal3DIdent



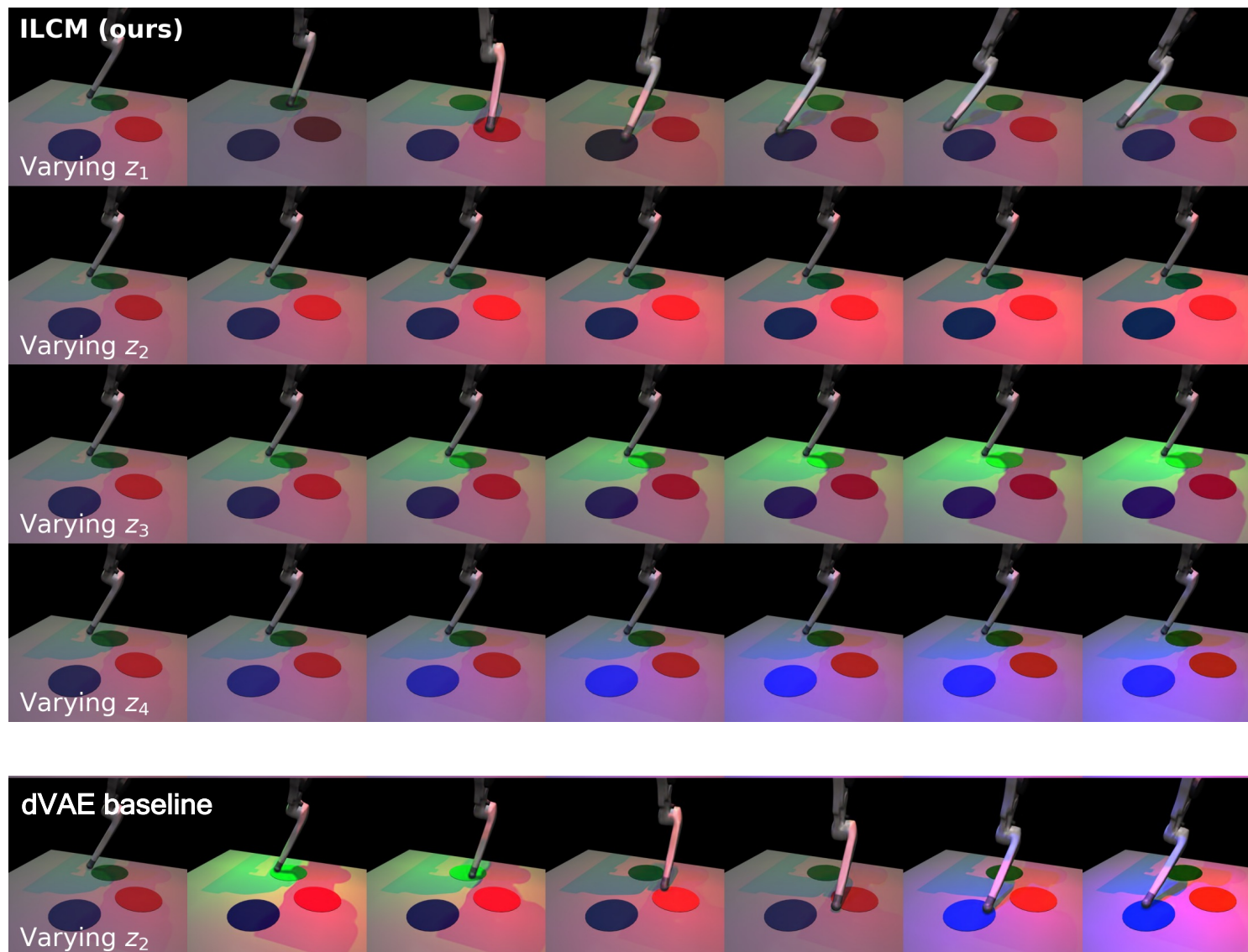
CausalCircuit

# CausalCircuit

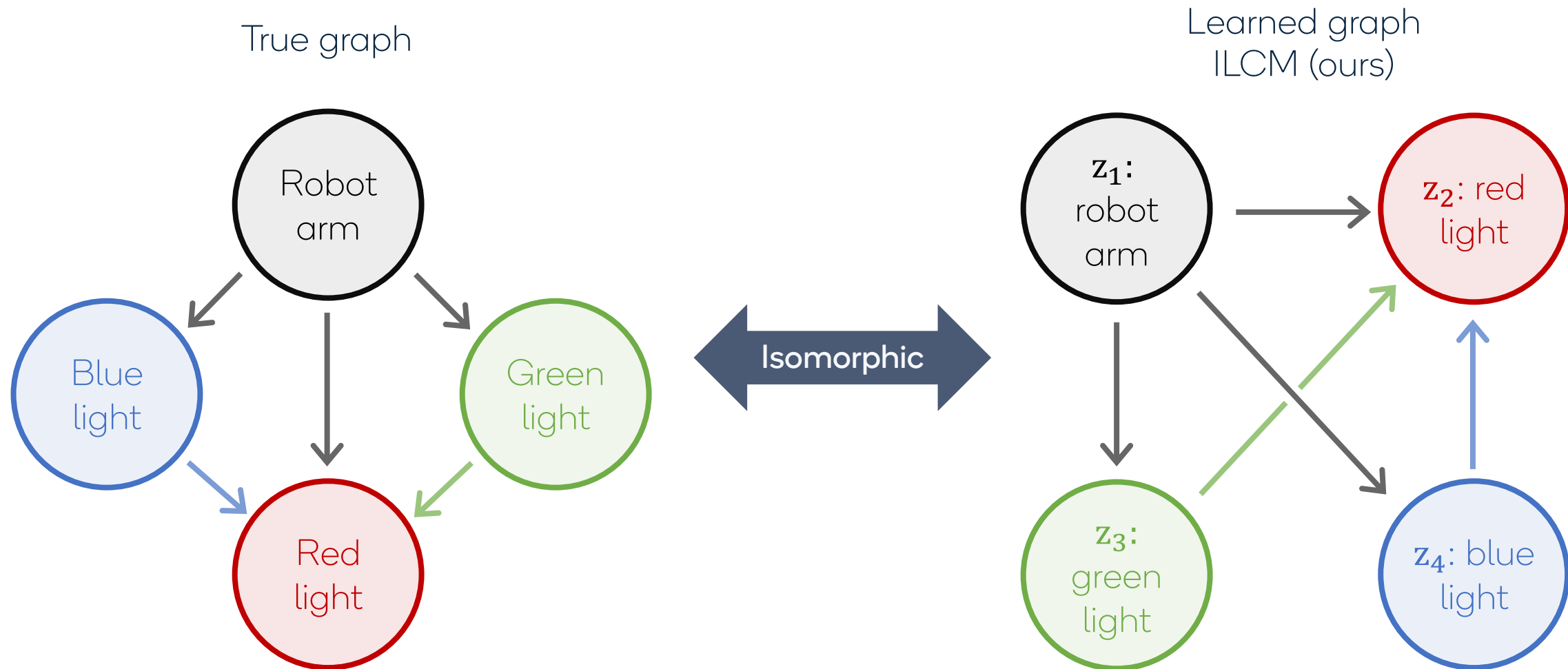
- **New dataset** with more intuitive causal structure
- **Robot arm interacts with touch-sensitive lights, which are connected with a circuit**
  - Robot arm movement based on inverse kinematic model
  - Physics + rendering with MuJoCo
  - 4 continuous causal variables: robot arm restricted to 1D arc + 3 light states
  - 512x512 images from fixed camera position
- ILCMs are trained on pre- and post-intervention data



# LCMs **disentangle** the causal variables



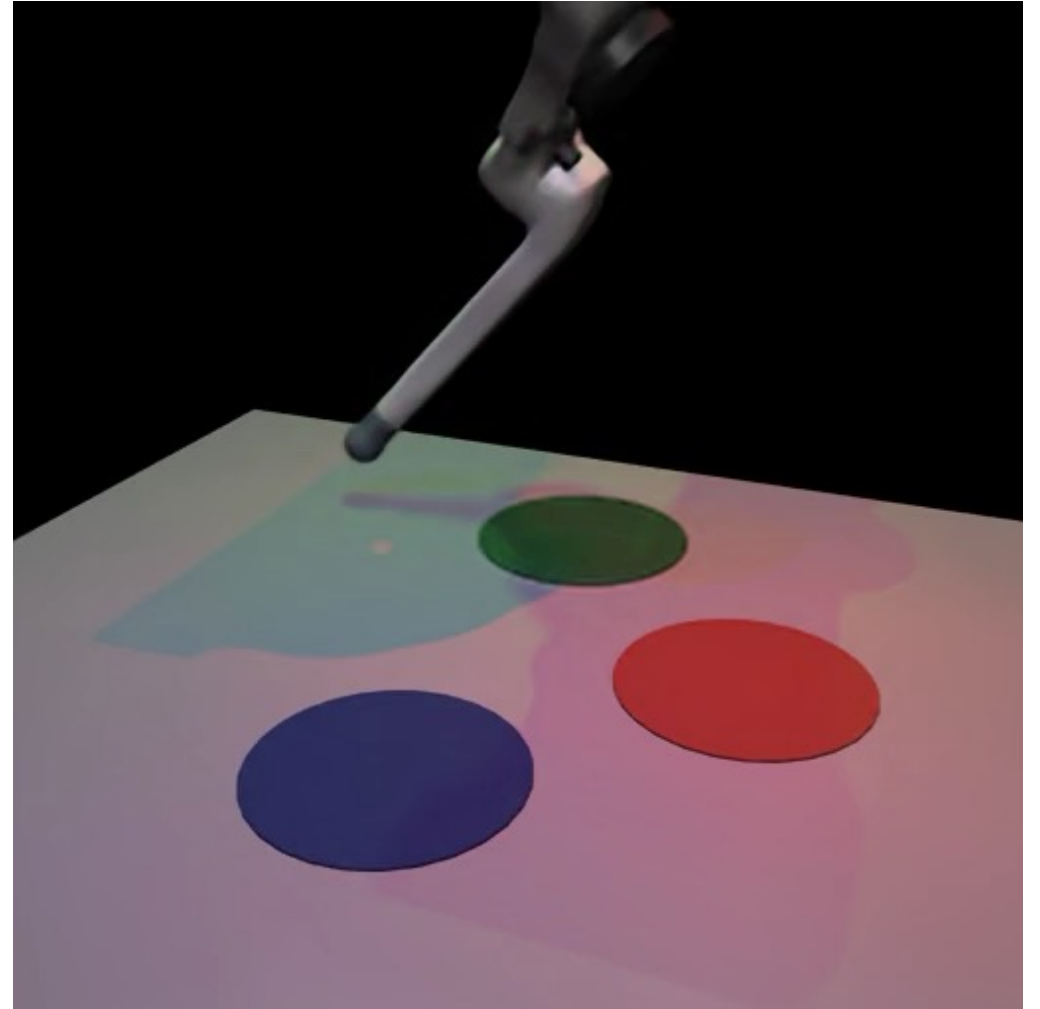
# LCMs learn the **correct** graph





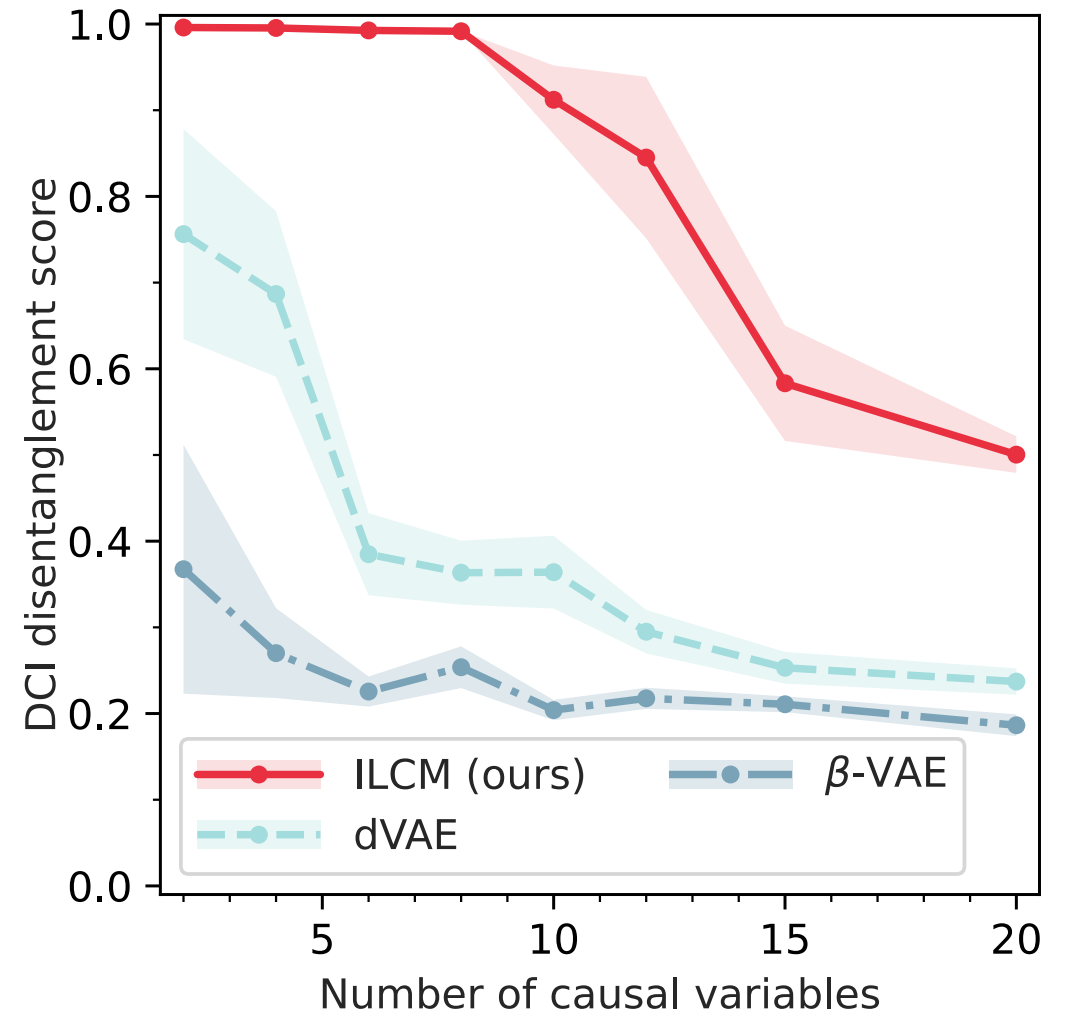
ILCMs let us **reason causally**

ILCM samples, **intervening** on a single latent  
(including causal effects)



# Do ILCMs **scale**?

- **Toy experiment:**
  - $n$  causal variables
  - linear causal effects
  - $SO(n)$  decoder
- ILCM results **robust up to ~10 variables** without additional tuning



Outlook

**Towards useful  
causal representation learning**

# A long way to go

Where we are

**Identifiability theorems**

**Pre- & post-intervention data**

**God-given interventions**

**Fixed causal variables**

**Strict DAG-based causality**

**Toy experiments** (up to  $O(10)$  variables)

Where we need to get

**Demonstrate usefulness** on downstream tasks

**Realistic data regimes:**  
observational & interventional data,  
video data, ...

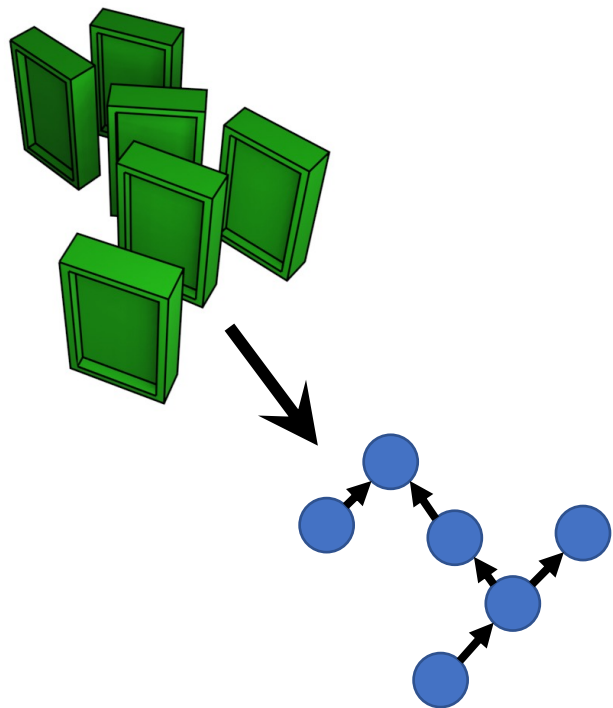
**Learning intervention policies**

**Variable scene composition**

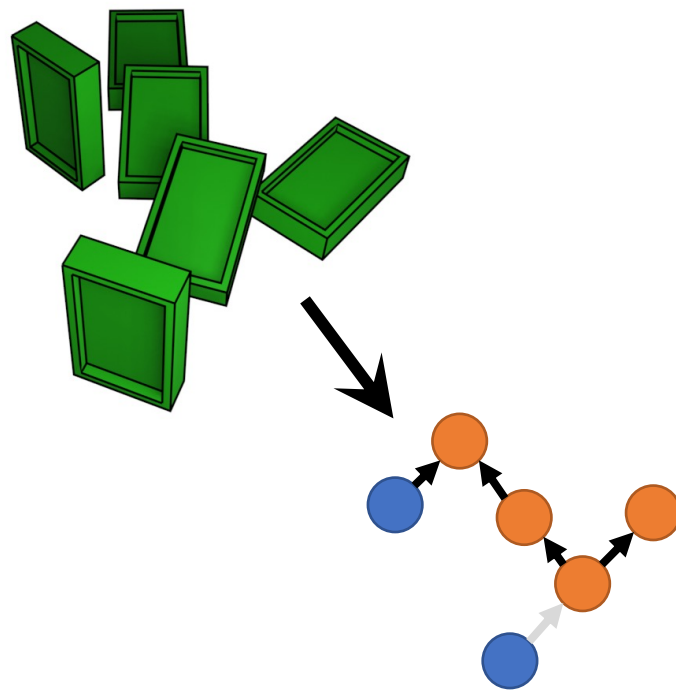
**Weaker relational structures**

**Realistic experiments**

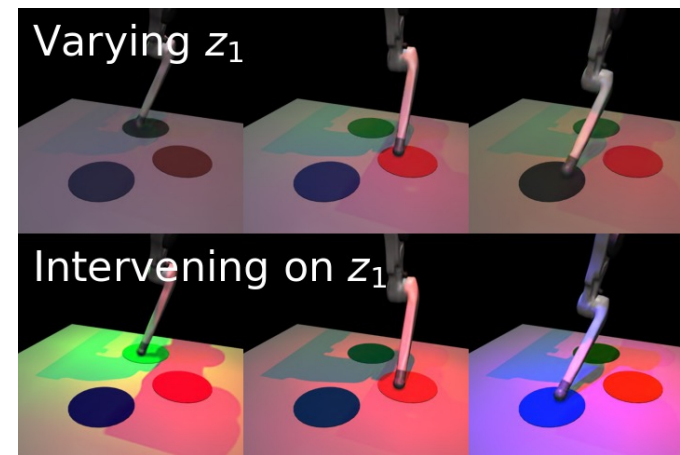




Can we **learn causal variables & causal structure from pixels**, without labels?



We prove: this is possible with **weak supervision**, when observing effects of interventions



In practice, **implicit latent causal models** can identify the causal structure in image datasets

## Weakly supervised causal representation learning

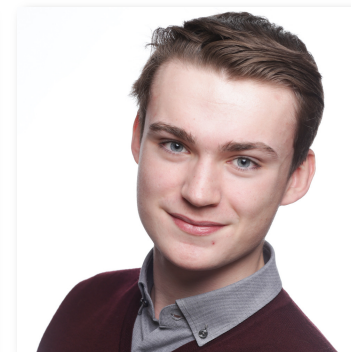
JB, Pim de Haan, Phillip Lippe, Taco Cohen

NeurIPS 2022

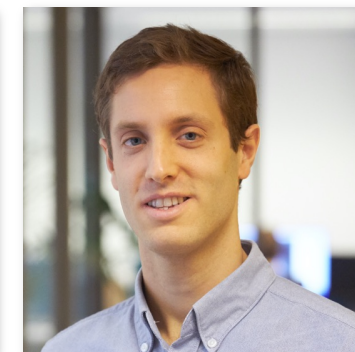
[arXiv:2203.16437](#)



Pim de Haan



Phillip Lippe



Taco Cohen

## Towards causal representation learning

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer,  
Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, Yoshua Bengio  
IEEE Advances in Machine Learning and Deep Neural Networks 2021

[arXiv:2102.11107](#)

## Weakly-supervised disentanglement without compromises

Francesco Locatello, Ben Poole, Gunnar Rätsch, Bernhard Schölkopf,  
Olivier Bachem, Michael Tschannen

ICML 2020

[arXiv:2002.02886](#)

## Self-supervised learning with data augmentations provably isolates content from style

Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel,  
Bernhard Schölkopf, Michel Besserve, Francesco Locatello  
NeurIPS 2021

[arXiv:2106.04619](#)

## CITRIS: Causal identifiability from temporal intervened sequences

Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco  
Cohen, Efstratios Gavves

ICML 2022

[arXiv:2202.03169](#)

# Thank you



Follow us on:     

For more information, visit us at:

[qualcomm.com](http://qualcomm.com) & [qualcomm.com/blog](http://qualcomm.com/blog)

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

©2018-2022 Qualcomm Technologies, Inc. and/or its affiliated companies. All Rights Reserved.

Qualcomm is a trademark or registered trademark of Qualcomm Incorporated. Other products and brand names may be trademarks or registered trademarks of their respective owners.

References in this presentation to "Qualcomm" may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes our licensing business, QTL, and the vast majority of our patent portfolio. Qualcomm Technologies, Inc., a subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of our engineering, research and development functions, and substantially all of our products and services businesses, including our QCT semiconductor business.