Reframing attention as a reinforcement learning problem for causal discovery

Turan Orujlu[1,2,3], **Christian Gumbsch**[4], **Martin V. Butz**[2], **Charley M. Wu**[1,3,5,6]
turan.orujlu@tuebingen.mpg.de

[1]**Human and Machine Cognition Lab, University of Tübingen, Germany**
[2]**Neuro-Cognitive Modeling Group, University of Tübingen, Germany**
[3]**Max Planck Institute for Biological Cybernetics, Germany**
[4]**VISLab, University of Amsterdam, Netherlands**
[5]**Centre for Cognitive Science, Technical University of Darmstadt, Germany**
[6]**Hessian.AI, Germany**

Abstract

Formal frameworks of causality have operated largely parallel to modern trends in deep reinforcement learning (RL). However, there has been a revival of interest in formally grounding the representations learned by neural networks in causal concepts. Yet, most attempts at neural models of causality assume static causal graphs and ignore the dynamic nature of causal interactions. In this work, we introduce Causal Process framework as a novel theory for representing dynamic hypotheses about causal structure. Furthermore, we present Causal Process Model as an implementation of this framework. This allows us to reformulate the attention mechanism popularized by Transformer networks within an RL setting with the goal to infer interpretable causal processes from visual observations. Here, causal inference corresponds to constructing a causal graph hypothesis which itself becomes an RL task nested within the original RL problem. To create an instance of such hypothesis, we employ RL agents. These agents establish links between units similar to the original Transformer attention mechanism. We demonstrate the effectiveness of our approach in an RL environment where we outperform current alternatives in causal representation learning and agent performance, and uniquely recover graphs of dynamic causal processes.

1 Introduction

Causality plays a fundamental role in building intelligent systems capable of physical reasoning (Gerstenberg et al., 2020). Explicitly modeling causal relationships is increasingly recognized to be crucial for developing robust, generalizable, and interpretable neural network models capable of accurate prediction and effective intervention (Xia et al., 2021). Despite their black-box nature, models such as transformers have demonstrated surprising capacity for causal reasoning (Nichani et al., 2024; Shou et al., 2023; Melnychuk et al., 2022; Dettki et al., 2025). One explanation posits that this is possible due to the attention mechanism forming implicit causal edges between tokens (Vaswani et al., 2017; Rohekar et al., 2023). However, recent work has highlighted a phenomenon known as *over-squashing*, in which the attention mechanism (and related message-passing mechanisms in Graph Neural Networks) loses sensitivity to individual tokens or nodes (Barbero et al., 2024a; Alon & Yahav, 2021; Barbero et al., 2024b; Giovanni et al., 2023; 2024; Topping et al., 2022; Scarselli et al., 2009; Battaglia et al., 2018). This compression of information in transformer models can sever causal chains, thus limiting the effectiveness of causal inference.

In contrast, graphical causal models, such as Pearl's Structural Causal Models (SCMs; Pearl, 2009), explicitly encode causal relationships and thus preserve perfect causal connectivity by design. Yet, a key challenge for SCMs is *causal discovery* (Schölkopf et al., 2021): inferring the causal graph from data. Most existing approaches assume access to a complete dataset and construct a static causal graph. This assumption is misaligned with the nature of physical environments, where causal influence is typically local in space and sparse in time (Pitis et al., 2020; Seitzer et al., 2021; Gumbsch et al., 2021; Lange & Kording, 2025). For instance, objects may only interact upon contact. Recent work has therefore emphasized the importance of *local causal models* (Pitis et al., 2020; Seitzer et al., 2021; Urpí et al., 2024; Lei et al., 2024; Willig et al., 2025) that explain causal connections through the sparsest possible graph changing dynamically over time.

Our work aims to bridge these areas by proposing a novel causal framework tailored to capture the dynamics of physical object interactions. We propose **Causal Process Models** (CPMs) which **reinterprets attention as a reinforcement learning problem**. Instead of computing soft attention, CPMs replace the transformer's attention mechanism with all-or-nothing connections determined by a reinforcement learning (RL) agent. While standard transformers process inputs through a fixed stack of blocks in a fixed order, CPMs introduce a dynamic architecture in which blocks can be conditionally selected, skipped, or reused. This allows the model to adaptively control both structure and depth based on the input, enabling more efficient and interpretable causal reasoning.

Our novel causal framework is designed specifically for modeling the dynamics of physical object interactions, aiming to synthesize the formal rigor of *static dependency* theories, e.g. Pearl's do-calculus (Pearl, 2009), with the intuitive strengths of *process-based* accounts (Russell, 1948; Salmon, 1984; Skyrms, 1981; Dowe, 2000, see Appendix A). Our approach explicitly addresses the limitations of Pearlian SCMs by enabling the construction of sparse, time-varying causal graphs that reflect only the active interactions between objects. When modeling two colliding balls for instance, our framework only instantiates a direct causal link between the balls upon contact, for the transfer of momentum, while leaving them causally disconnected otherwise. This yields a computationally efficient model, scaling with actual rather than potential interactions, and one that is highly interpretable since the causal graph mirrors intuitive physical processes.

Our main contribution are as follows

1. We formalize a framework for local causal modeling in physical environments.
2. We implement this framework in a neural architecture that dynamically infers sparse, time-varying causal graphs by reinterpreting attention as a decision making problem.
3. We apply our model to a canonical physical interaction scenario demonstrating superior performance and interpretability compared to densely connected models.

2 Causal Process Framework

Pearl's structural causal models (SCMs) and do-calculus (Pearl, 2009) provide a powerful foundation for causal reasoning. However, SCMs, by design, cannot be applied to dynamic physical systems requiring object-centric representations and real-time causal interactions. Prior approaches (Buesing et al., 2019; Gasse et al., 2023) have attempted to bridge model-based RL and causality by representing the full Markov Decision Process (MDP) state s^t using a single node and modeling actions as direct interventions in a static causal graph. However, this approach is limited because it circumvents the problem of inferring the causal structure that generates the underlying environment dynamics (i.e., the causal context; Butz et al., 2025), and focuses only on the causal implications of action sequences.

2.1 Causal Process Models (CPMs)

Here, we adopt an *object-centric factorization* of states, where each *object* in a scene as well as the *forces* causally influecing them are modeled as nodes (see Appendix B for details). Let $\mathcal{F}$ and $\mathcal{O}$ represent the sets of force nodes and object nodes, respectively. Let's denote the power set of all

possible edges between forces $F \in \mathcal{F}$ and objects $O \in \mathcal{O}$ as $\mathcal{E}(\mathcal{F}, \mathcal{O})$. For simplicity, we assume the existence of only one type of object and force nodes. In the future, our framework will be extended to multiple types.

We introduce two types of *controller functions* that control the construction of a causal subgraph $G_{\mathcal{O}^t:\mathcal{O}^{t+1}}$ defining the causal chain of object-to-force and force-to-object connections, respectively. The *interaction scope controllers*, $\rho^t_{\mathcal{O}}$, are expressed as a probability distribution over edge subsets $\mathcal{E}(\mathcal{O}^t, \mathcal{F}^t)$ conditioned on $\mathcal{O}^t$,

$$\begin{aligned} \rho^t_{\mathcal{O}} : \mathcal{E}\left(\mathcal{O},^t \mathcal{F}^t\right) &\to [0,1], \\ J^t &\mapsto \mathbb{P}\left[G_{\mathcal{O}^t:\mathcal{F}^t} = \left(\left\{O^t_n, F^t_m\right\}_{(m,n)\in M\times N}, \left\{E\left(O^t_i, F^t_j\right)\right\}_{(i,j)\in J^t}\right)\right]. \end{aligned} \tag{1}$$

The *effect attribution controllers* , $\rho^t_{\mathcal{O}\leftrightarrow\mathcal{F}}$ describe a distribution over edge subsets $\mathcal{E}\left(\mathcal{F}^t, \mathcal{O}^{t+1}\right)$ and are conditioned on both $\mathcal{O}^t$ and $\mathcal{F}^t$:

$$\begin{aligned} \rho^t_{\mathcal{O}\leftrightarrow\mathcal{F}} : \mathcal{E}\left(\mathcal{F},^t \mathcal{O}^{t+1}\right) &\to [0,1], \\ I^t &\mapsto \mathbb{P}\left[G_{\mathcal{F}^t:\mathcal{O}^{t+1}} = \left(\left\{F^t_m, O^{t+1}_n\right\}_{(m,n)\in M\times N}, \left\{E\left(F^t_j, O^{t+1}_i\right)\right\}_{(j,i)\in I^t}\right)\right]. \end{aligned} \tag{2}$$

In the equations above, $E\left(O^t_i, F^t_j\right)$ and $E\left(F^t_j, O^{t+1}_i\right)$ represent directed edges from O^t_i to F^t_j and from F^t_j to O^{t+1}_i, respectively. The sets of index tuples $J^t \subset N \times M$ and $I^t \subset M \times N$ identify these edges, $G_{\mathcal{O}^t:\mathcal{F}^t} := \left(G^{\mathcal{V}}_{\mathcal{O}^t:\mathcal{F}^t}, G^{\mathcal{E}}_{\mathcal{O}^t:\mathcal{F}^t}\right)$ and $G_{\mathcal{F}^t:\mathcal{O}^{t+1}} := \left(G^{\mathcal{V}}_{\mathcal{F}^t:\mathcal{O}^{t+1}}, G^{\mathcal{E}}_{\mathcal{F}^t:\mathcal{O}^{t+1}}\right)$ stand for the random variables describing the causal subgraphs with nodes $G^{\mathcal{V}}_{\mathcal{O}^t:\mathcal{F}^t}$, edges $G^{\mathcal{E}}_{\mathcal{O}^t:\mathcal{F}^t}$ and nodes $G^{\mathcal{V}}_{\mathcal{F}^t:\mathcal{O}^{t+1}}$, edges $G^{\mathcal{E}}_{\mathcal{F}^t:\mathcal{O}^{t+1}}$ respectively.

As the name suggests, interaction scope controllers decide the causal context for interactions, i.e., they decide which objects are currently relevant. The effect attribution controllers choose to which force node to relay the context information. With the help of these controllers, we define the CPM by specifying how each node is updated over time:

$$\begin{aligned} F^t_j &:= f_F\left(F^{t-1}_j, \left\{O^{t-1}_i\right\}_{i|(i,j)\in J^{t-1}}\right) \qquad \text{s.t. } J^{t-1} \sim \rho^{t-1}_{\mathcal{O}}, \\ O^t_i &:= f_O\left(O^{t-1}_i, \left\{F^t_j\right\}_{j|(j,i)\in I^{t-1}}\right) \qquad \text{s.t. } I^{t-1} \sim \rho^{t-1}_{\mathcal{O}\leftrightarrow\mathcal{F}}. \end{aligned} \tag{3}$$

Thus, update functions f_F and f_O are force- and object-specific (respectively) and invariant to the number of inputs (i.e., size of the parent node set). When interventions occur at time step $\tilde{t}$, they introduce new force nodes, denoted as $\text{act}(F^{\tilde{t}}_*)$ representing externally applied actions. At this point, the subgraph corresponding to time steps $t \geq \tilde{t}$ gets recomputed following Algorithm S1.

2.2 Inductive biases

To constrain complexity and utilize prior knowledge, we introduce **two inductive biases**. First, we **restrict each force node to connect to exactly two different object nodes**. This corresponds to the assumption that typically not more than two objects interact at a certain time step. This restriction can be lifted later to generalize to hypergraphs for more complicated systems (e.g., 3-body problem). Second, we **enforce a mirroring constraint between force nodes and object nodes** to ensure coherence in causal attribution and prevent nonsensical unidirectional relationships. Intuitively, at time $t+1$, a force must influence exactly one of the objects it has taken as input at time t. Similarly, if an object node has been affected by a force node at time $t+1$, it must have affected that same force node in the previous time step t (see Fig. 1 and Appendix C.1 for a formal definition).

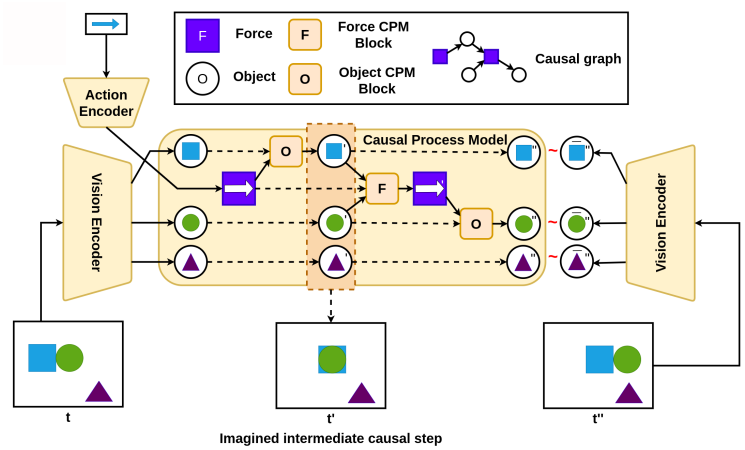


Figure 1: Our model has three components: a vision encoder, an action encoder, and the transition function. The transition function is an implementation of a *Causal Process Model*. The state is factorized into distinct object representations, actions are mapped to force representations, and the directed edges are causal. The model is trained via contrastive loss (Eq. 8).

3 Model

We base our model implementation on the Structured World Model (C-SWM; Kipf et al., 2020). The model consists of an *object-centric vision encoder*, an *action encoder*, and a *transition function*. We keep the structure of the vision and action encoders intact, but modify the transition function.

The *vision encoder* is a CNN-based object extractor E_{ext}, operating directly on images and outputting I feature maps. Each feature map $m_i^t = [E_{\text{ext}}(s^t)]_i$ acts as an object mask where $[\dots]_i$ is the selection of the i^{th} feature map. An MLP-based object encoder E_{enc} with shared weights across objects maps the flattened feature map m_i^t to object latent representation: $O_i^t = E_{\text{enc}}(m_i^t)$. Additionally, an MLP-based *action encoder* maps action a^t to force latent representation: $F^t = A(a^t)$.

Next we introduce our new transition function (Sec. 3.1) before detailing how construct the causal graph on the fly using reinforcement learning (Sec. 3.2).

3.1 Causal Transition Function

Our main innovation is the *transition function* as a neural network implementation of a CPM. We use two feedforward neural networks, $f_F(\dots;\theta_F)$ and $f_O(\dots;\theta_O)$, shared by all the force and object nodes respectively. The force F_j^t and object O_i^t vectors processed by these neural networks come with inductive biases. Based on mutability, causal relevance, and control relevance, the force vector F_j^t is divided into $2^3 - 1 = 7$ (all possible combinations except the case with a sub-vector carrying none of these features) sub-vectors of equal size d_F where a sub-vector's identity determines how it is being affected by the neural networks:

$$F_j^t := \left[\begin{bmatrix} M & C & K \\ \bullet & \bullet & \bullet \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \bullet & \bullet & \circ \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \bullet & \circ & \bullet \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \bullet & \circ & \circ \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \circ & \bullet & \bullet \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \circ & \bullet & \circ \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \circ & \circ & \bullet \end{bmatrix}_{F_j}^t \right]. \tag{4}$$

In Eq. 4, M stands for *mutability*, C for *causal* relevance, K for *control* relevance, and $\bullet$ indicates the presence of a feature, while $\circ$ is the absence. We are going to use the following shorthand notation:

$$\begin{bmatrix} M \\ \bullet \end{bmatrix}_{F_j}^t := \left[\begin{bmatrix} M & C & K \\ \bullet & \bullet & \bullet \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \bullet & \bullet & \circ \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \bullet & \circ & \bullet \end{bmatrix}_{F_j}^t ; \begin{bmatrix} M & C & K \\ \bullet & \circ & \circ \end{bmatrix}_{F_j}^t \right].$$

The object vector O_i^t is divided into 8 subvectors with the first 7 being similar to that of F_j^t and an additional $\begin{bmatrix} M & C & K \\ \circ & \circ & \circ \end{bmatrix}_{O_i}^t$ that stands for the part of the object representation that does not change, and is neither causally, nor control relevant. The goal here is to enforce inductive biases that will result in learned semantic encoding; for example, if the force F represents collision between two objects O_1 and O_2, the masses of these objects will be learned to be causally relevant but immutable, belonging to $\begin{bmatrix} M & C & K \\ \circ & \bullet & \circ \end{bmatrix}_{O_i}^t$. However, their velocities will be learned to be both mutable and causally relevant belonging to $\begin{bmatrix} M & C & K \\ \bullet & \bullet & \circ \end{bmatrix}_{O_i}^t$. Additionally, their light reflectance will be learned to be mutable but not causally relevant for the collision momentum belonging to $\begin{bmatrix} M & C & K \\ \bullet & \circ & \circ \end{bmatrix}_{O_i}^t$. See Appendix C.2 for the further constraints we impose on vector representations.

The actual implementation of the causal process blocks $f_O(\cdots \mid \theta_O)$ and $f_F(\cdots \mid \theta_F)$ is quite similar to that of transformer blocks, but with the attention mechanism (Vaswani et al., 2017) replaced by indices of the chosen force and object nodes (tokens in transformers). See Appendix C.3 and Fig. S1 for more details.

3.2 Causal Controller

The main proposal of our model is that we perform graph construction through sequential decision making using the interaction scope and effect attribution controllers. Thus, we treat causal discovery as a multi-agent RL problem, where one agent determines the scope of interacting objects ($\pi_O(\mathcal{O}^t) := \rho_{\mathcal{O}}^t$) and another agent determines how the force effects are attributed ($\pi_{O\leftrightarrow F}\left(\mathcal{O}^t, \mathcal{F}^{t+1}\right) := \rho_{\mathcal{O}\leftrightarrow\mathcal{F}}^t$).

More specifically, for the computations above, we assume the chosen indices I^t and J^t are provided by the agents $\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}$ and $\pi_{\mathcal{O}}$. The two agents alternate outputting an action. An action taken by $\pi_{\mathcal{O}}$ corresponds to two edge additions $E\left(O_i^t, F^{t+1}\right), E\left(O_j^t, F^{t+1}\right), i \neq j$ to the graph. Whereas an action taken by $\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}$ results in one edge addition $E\left(F^t, O_i^t\right)$ (see Fig. S2). The index set I^t is sampled using the policy of the agent $\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}$:

$$I^t \sim \pi_{\mathcal{O}\leftrightarrow\mathcal{F}}\left(I^t \mid G^t, \psi_{\mathcal{O}\leftrightarrow\mathcal{F}}\right) = \text{softmax}\left(\left(\left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_F^t W_{\text{agent}}^{\mathcal{F}}\right)\left(\left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_O^t W_{\text{ctrl}}^{\mathcal{F}}\right)^T / d\right), \tag{5}$$

where $\psi_{\mathcal{O}\leftrightarrow\mathcal{F}} := \left\{W_{\text{agent}}^{\mathcal{F}}, W_{\text{ctrl}}^{\mathcal{F}}\right\}$ and d is the dimension of $W_{\text{agent}}^{\mathcal{F}}$'s codomain, $W_{\text{agent}}^{\mathcal{F}} \in \mathbb{R}^{4d_F \times d}$. $J_{\mathcal{O}}^t$, on the other hand, is sampled using the policy of the agent $\pi_{\mathcal{O}}$:

$$J^t \sim \pi_{\mathcal{O}}\left(J^t \mid G^t, \psi_{\mathcal{O}}\right) = \text{softmax}\left(\text{triu}\left(\left(\left[\left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_O^t, \vec{\emptyset}\right] W_{\text{agent}}^{\mathcal{O}}\right)\left(\left[\left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_O^t, \vec{\emptyset}\right] W_{\text{ctrl}}^{\mathcal{O}}\right)^T\right) / d\right), \tag{6}$$

where $\psi_{\mathcal{O}} := \left\{W_{\text{agent}}^{\mathcal{O}}, W_{\text{ctrl}}^{\mathcal{O}}\right\}$, $\vec{\emptyset}$ is a dummy variable indicating no-choice, and triu is an operator picking out the upper triangular segment of a matrix (excluding its diagonal), and d is the same as above.

We then define separate reward functions for $\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}$ and $\pi_{\mathcal{O}}$:

$$\begin{aligned} R_{\mathcal{O}\leftrightarrow\mathcal{F}}\left(G^t, G^{t+1} \mid \theta_{R_{\mathcal{O}\leftrightarrow\mathcal{F}}}\right) &= \text{MLP}\left(\left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_F^t, \left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_O^{t-1}, E\left(F^t, O_i^t\right), \left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_O^t \middle| \theta_{R_{\mathcal{O}\leftrightarrow\mathcal{F}}}\right), \\ R_{\mathcal{O}}\left(G^t, G^{t+1} \mid \theta_{R_{\mathcal{O}}}\right) &= \text{MLP}\left(\left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_O^t, \left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_F^t, E\left(O_j^t, F^{t+1}\right), E\left(O_k^t, F^{t+1}\right), \left[\begin{smallmatrix} K \\ \bullet \end{smallmatrix}\right]_F^{t+1} \middle| \theta_{R_{\mathcal{O}}}\right). \end{aligned} \tag{7}$$

Note that this reward is a priori undefined and needs to be learned, making this an inverse reinforcement learning task.

4 Training

We use three-staged training for our model. In the first stage, we freeze the weights of $\pi_{\mathcal{O}}$ and $\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}$, and train the rest of the model assuming a complete causal graph with contrastive loss (Kipf et al., 2020):

$$\begin{aligned}\mathcal{L}_{\text{CPM}}\left(\Theta|s^t,a^t,s^{t+1},\tilde{s}^t,\mathcal{K}\right) = &\left\|\text{CPM}\left(\text{V}\left(s^t|\theta_V\right),\text{A}\left(a^t|\theta_A\right),\mathcal{K}|\theta_O,\theta_F\right)-\text{V}\left(s^{t+1}|\theta_V\right)\right\| \\ &+\max\left(0,\beta-\left\|\text{V}\left(\tilde{s}^t|\theta_V\right)-\text{V}\left(s^{t+1}|\theta_V\right)\right\|\right),\end{aligned} \tag{8}$$

where $\Theta=[\theta_V;\theta_A;\theta_O;\theta_F;\theta_{R_{\mathcal{O}\leftrightarrow\mathcal{F}}};\theta_{R_{\mathcal{O}}}]$, CPM is the Causal Process Model, V is a vision encoder, A is an action encoder, s^t is state, a^t is action, $\mathcal{K}$ is a layered multi-partite graph with adjacent layers forming complete bi-partite sub-graphs, and $\tilde{s}^t$ is a state sampled at random from the experience buffer to serve as a negative example for the contrastive loss. The first summand is the prediction error and the second summand is contrastive regularizer. The hinge margin β is set to be 1 as recommended by Kipf et al. (2020).

In the second stage, we flip the roles to train the parameters of $\pi_{\mathcal{O}}$ and $\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}$ using policy gradient (Williams, 1992), and freeze the rest of the model. The loss at this stage is defined as:

$$\begin{aligned}\mathcal{L}_\rho\left(\Psi\left|\left(G^t,I^t,J^t,G^{t+1}\right)_t,\Theta\right.\right) = &-\sum_t\log\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}\left(I^t|G^t,\psi_{\mathcal{O}\leftrightarrow\mathcal{F}}\right)\sum_\tau\gamma^\tau R_{\mathcal{O}\leftrightarrow\mathcal{F}}\left(G^\tau,G^{\tau+1}|\theta_{R_{\mathcal{O}\leftrightarrow\mathcal{F}}}\right) \\ &-\sum_t\log\pi_{\mathcal{O}}\left(J^t|G^t,\psi_{\mathcal{O}}\right)\sum_\tau\gamma^\tau R_{\mathcal{O}}\left(G^\tau,G^{\tau+1}|\theta_{R_{\mathcal{O}}}\right),\end{aligned} \tag{9}$$

where $\Psi=[\psi_{\mathcal{O}\leftrightarrow\mathcal{F}};\psi_{\mathcal{O}}]$.

In the third stage, we employ iterative training of the CPM and agents $\pi_{\mathcal{O}}$, $\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}$ where we minimize $\mathcal{L}_{\text{CPM}}$ and $\mathcal{L}_\rho$ in an alternating fashion. However, we add regularization terms to $\mathcal{L}_{\text{CPM}}$ to account for reward learning:

$$\begin{aligned}&\mathcal{L}^\mu_{\text{CPM}}\left(\Theta\left|s^t,a^t,s^{t+1},\tilde{s}^t,\left(G^\tau,I^\tau,J^\tau,G^{\tau+1}\right)_\tau,\Psi\right.\right)=\mathcal{L}_{\text{CPM}}\left(\Theta\left|s^t,a^t,s^{t+1},\tilde{s}^t,\left(I^\tau,J^\tau\right)_\tau\right.\right)+ \\ &\sum_\tau\mu\left\|R_{\mathcal{O}\leftrightarrow\mathcal{F}}\left(G^\tau,G^{\tau+1}|\theta_{R_{\mathcal{O}\leftrightarrow\mathcal{F}}}\right)-\left(\log\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}\left(I^\tau|G^\tau,\psi_{\mathcal{O}\leftrightarrow\mathcal{F}}\right)-\gamma\max_{I^{\tau+1}}\log\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}\left(I^{\tau+1}|G^{\tau+1},\psi_{\mathcal{O}\leftrightarrow\mathcal{F}}\right)\right)\right\|+ \\ &\sum_\tau\mu\left\|R_{\mathcal{O}}\left(G^\tau,G^{\tau+1}|\theta_{R_{\mathcal{O}}}\right)-\left(\log\pi_{\mathcal{O}}\left(J^\tau|G^\tau,\psi_{\mathcal{O}}\right)-\gamma\max_{J^{\tau+1}}\log\pi_{\mathcal{O}}\left(J^{\tau+1}|G^{\tau+1},\psi_{\mathcal{O}}\right)\right)\right\|,\end{aligned} \tag{10}$$

where we assume $\log\pi_{\mathcal{O}\leftrightarrow\mathcal{F}}\left(I^t|G^t,\psi_{\mathcal{O}\leftrightarrow\mathcal{F}}\right)\propto Q\left(G^t,I^t\right)$ and $\log\pi_{\mathcal{O}}\left(J^t|G^t,\psi_{\mathcal{O}}\right)\propto Q\left(G^t,J^t\right)$. The regularization terms are a reformulation of the converged Q-learning update (Watkins & Dayan, 1992). We set $\mu=0.1$, $\gamma=0.9$ in the above loss functions.

5 Experiments

We hypothesize that our model outperforms models that assume dense causal graphs to capture physical interactions in: 1) longer prediction horizons; 2) test-time generalization across unobservable properties; 3) robustness with regards to the number of objects in the scene; 4) solving downstream

tasks. We use the *physics environment* designed by Ke et al. (2021) to empirically answer these questions (Fig. 2 top). The environment consists of different objects colored according to their weights. The only force in this environment is pushing (double-pushes are not allowed) and only heavier objects can push lighter ones. The environment has two settings: an *observed* setting (Fig. 2a) where weight corresponds to the intensity of a particular color and an *unobserved* setting (Fig. 2b) where different colors did not systematically map to different weights.

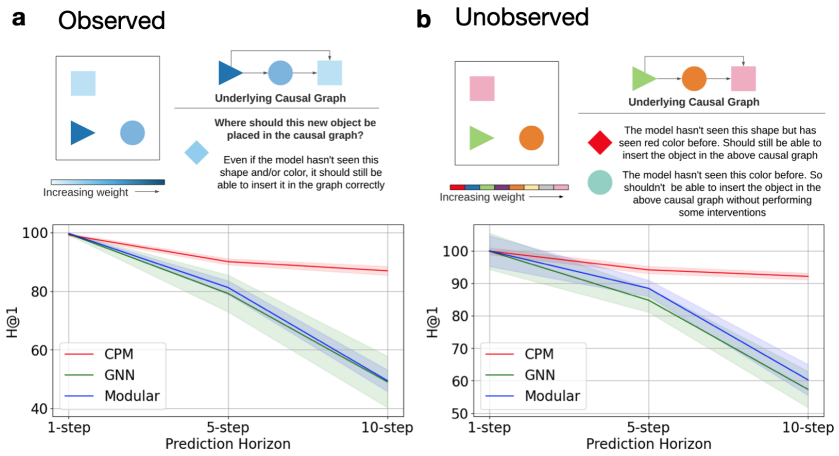


Figure 2: *Prediction results for a synthetic physics environment* in a) observed and b) unobserved settings (Ke et al., 2021) **Top:** Description of the task. **Bottom:** Prediction metric vs prediction horizon for 5 objects (average of top 8 out of 10 seeds).

5.1 Comparison Baselines

We compare our model against 2 baselines, a graph neural network (GNN) (Scarselli et al., 2009) and a modular network which has a separate MLP to model each object's dynamics.

5.2 Prediction Metrics

To investigate robustness towards the length of prediction horizons, we trained the model to make 1-step predictions in the *Observed* setting with 5 objects and then tested for 5 and 10 steps (Fig. 2a bottom). We used Hits at Rank 1 (H@1) to measure model performance as an all-or-nothing metric measuring how often the rank of the predicted representation was 1 when ranked against all reference state representations (see Figs. S6-S8 for comparison to other metrics). Here, our model consistently outperformed the baseline models, with the gap increasing over longer time horizons.

Next, to estimate the test-time generalization across unobservable properties, we trained our model in the *Unobserved* setting where generalization at test time is harder due to previously unseen weights. Again, our model consistently outperformed the baselines displaying capacity to generalize also in this domain (Fig. 2b bottom; see Figs. S3-S8 for more results).

Additionally, we tested robustness with regards to the number of objects in the scene. We trained our model to make 1-step predictions in the Observed setting with 3-7 objects and then tested for 5 steps (Fig. 3a). Here as well, our model outperformed the baselines in all but 3-object setting.

5.3 Downstream RL tasks

To make sure the above metrics overlap with the learned model's usefulness for downstream tasks, we also tested our CPM's capacity to serve as a world model for a model-based RL agent. The agent's task was to move an object to a certain location where failure resulted in negative reward. In the Observed setting, our model outperformed the baselines for all settings except for the 3-object

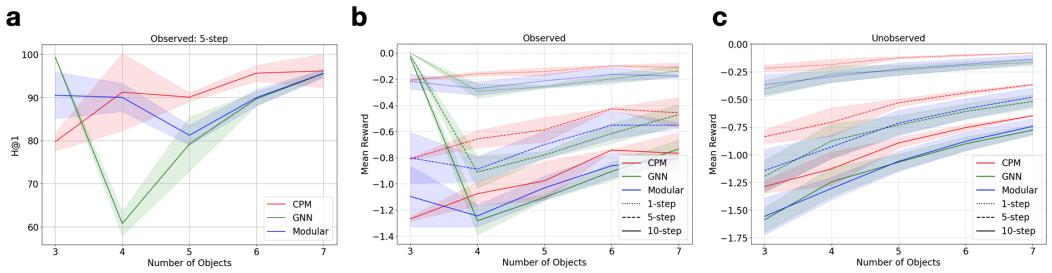


Figure 3: *Prediction and downstream RL results over number of objects* **a:** Prediction metric vs number of objects for 5-steps. **b-c:** Mean reward vs number of objects. All results are the average of top 8 out of 10 seeds

setting and 7-object 10-step prediction setting (Fig. 3b). In the Unobserved setting, our model was consistently better than the baselines (Fig. 3c).

6 Discussion

In this paper, we introduced the Causal Process Framework, a novel approach for modeling the dynamics of physical object interactions. Our key contribution is the Causal Process Model (CPM), which implements this framework by reframing the attention mechanism as a reinforcement learning problem. Instead of the soft, dense connections typical of Transformers, our model employs RL agents to dynamically construct sparse, time-varying causal graphs. Our experiments in a simulated physics environment show that this approach not only improves prediction accuracy and downstream task performance compared to GNN and modular baselines, but also excels in generalization and scalability.

The superior performance of our model, particularly over longer prediction horizons and with a varying number of objects, lends strong support to our central hypothesis. We argue that by explicitly modeling only active causal links, the CPM avoids the pitfalls of dense message-passing architectures. Our discrete, "all-or-nothing" connections, determined by a goal-oriented RL agent, preserve the salience of individual interactions. This leads to more robust and precise world models, which proved crucial for the model-based RL agent's success in downstream tasks. Furthermore, the model's ability to generalize to unobserved object properties suggests that it learns an underlying model of physical dynamics rather than memorizing superficial correlations.

Despite these promising results, the present work has several limitations that open clear avenues for future research. A crucial next step is to benchmark our CPM against a standard Transformer architecture. This comparison will directly test whether our explicit, learned causal graph provides tangible benefits over the implicit causal reasoning capabilities of soft attention mechanisms (Nichani et al., 2024). Another priority is to deepen the analysis of the learned representations. We will conduct experiments to decode the semantic content of the force and object sub-vectors (e.g., mutable, causal, and controllable components) to verify that our inductive biases are effective in fostering an interpretable internal structure. To fully validate our claims of causal discovery, we will also compare the inferred graphs against the ground-truth interaction graphs of the simulation, providing a quantitative measure of the model's ability to recover the true causal processes.

Looking forward, we intend to scale our model to more complex and realistic domains. A key challenge will be to extend the framework beyond pairwise interactions to handle hypergraphs, which are necessary for scenarios like the three-body problem or multi-object collisions. Ultimately, we aim to apply our CPM to high-dimensional, real-world settings, such as controlling a 3D robotic arm from visual input. Success in these environments would demonstrate the framework's potential to bridge the gap between formal causal reasoning and modern deep learning, paving the way for more robust, generalizable, and interpretable agents that can safely and effectively interact with the physical world.

A Related Work

A.1 Causal frameworks

One dominant approach to causal modeling stems from Pearl's framework of Structural Causal Models (SCMs) (Pearl, 2009), which represents causal relationships using directed acyclic graphs (DAGs). An SCM can be described as a tuple $\mathfrak{C} := (\mathbf{S}, \mathbb{P}(\mathbf{U}))$ where $\mathbb{P}$ is a distribution over the exogenous variables $\mathbf{U}$ (i.e., variables external to the system and not caused by any variable within it) and $\mathbf{S}$ is a collection of structural equations of the form:

$$V_i = f_{V_i}(\mathbf{Pa}_{V_i}, \mathbf{U}_{V_i}).$$

Each endogenous variable V_i is determined by a function of its parent variables $\mathbf{Pa}_{V_i}$ (i.e., other variables in the system that directly influence V_i) and its associated exogenous noise term $\mathbf{U}_{V_i}$.

While successful in many domains, standard SCMs encounter significant hurdles when applied to systems characterized by dynamic object interactions. Consider the simple scenario of two colliding balls. Representing this within a traditional SCM framework often requires specifying potential causal links between all properties of all objects at all relevant timescales. This leads to densely connected causal graphs, with the number of causal edges scaling quadratically with time, even when interactions are sparse in reality. Such dense representations suffer from high computational costs for inference and learning, and crucially, obscure the underlying causal structure, hindering interpretability. Thus, a core challenge is to adapt standard SCMs to dynamically represent only the relevant interactions as they occur, rather than needing to specify all potential dependencies.

Recognizing these limitations, other lines of research offer valuable perspectives, often aligning closely with *causal process theories* (Russell, 1948; Salmon, 1984; Skyrms, 1981; Dowe, 2000). Research in cognitive science, such as Gerstenberg et al. (2020)'s counterfactual simulation models, leverage simulation to assess causality and responsibility in physical events, capturing process-like intuitions. Furthermore, philosophical inquiries into causal processes by thinkers like Russell, Salmon, Skyrms, and Dowe provide rich conceptual foundations, distinguishing *causal* processes from *pseudo*-processes by focusing on mechanisms like causal lines (Russell, 1948), defining causality in ontological terms (Salmon, 1984), or using conserved quantities (Skyrms, 1981; Dowe, 2000), but typically lack the computational formalism required for direct implementation in ML systems.

A.2 Neural Causal Models

Previous attempts to reconcile deep learning with SCMs have resulted in Neural Causal Models (NCMs), which model f_{V_i} as feedforward neural nets parametrized by θ_{V_i} (Xia et al., 2021). This solution is not satisfactory for us since that implies training arbitrarily many feedforward neural networks for each node across time. Zecevic et al. (2021) have to tried to theoretically establish GNNs capacity to implement SCMs, but as such, they also assume static causal graph. Melnychuk et al. (2022) designed a Causal Transformer that can infer causality over time, yet lacks an explicit dynamic causal graph. This is due to the reliance on the potential outcomes framework, which is not a graph-based approach (Rubin, 1978; Robins & Hernan, 2008).

A.3 Causal Reinforcement Learning

Buesing et al. (2019) have tried to take advantage of the Pearlian causality framework by reformulating the MDP graph as an SCM using which they designed a counterfactually-guided policy search. A similar approach has been pursued by Gasse et al. (2023) in which they draw parallels between confounding variables and offline RL. Neither of these approaches considers the causal structure that generates the dynamics of the environment as suggested by Bareinboim et al. (2021).

B Object-centric Causal Dynamics

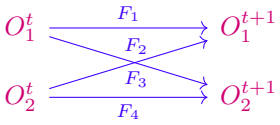
Consider two objects O_1 and O_2 (depicted in magenta) with a force F (depicted in violet) acting on them:

$$O_1 \xrightarrow{F} O_2$$

This recovers the familiar structure of a directed acyclic graphs (DAGs) from Pearl's causal formalism Pearl (2009). However, in physical interactions, such as in a collision, it is not always clear which object is the "cause" since both are affected simultaneously. A more intuitive representation would be a bidirectional edge:

$$O_1 \xleftrightarrow{F} O_2$$

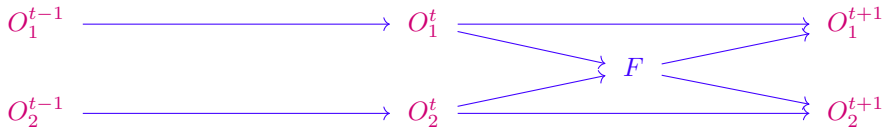
However, DAGs prohibit cycles and bidirectional edges. To resolve this, we introduce *temporal dynamics* which represent causal effects as unfolding over time rather than as a simultaneous influence. Thus, a collision between object O_1 and object O_2 yields forces F_2 and F_3 as emerging from the past state and influencing future object states:



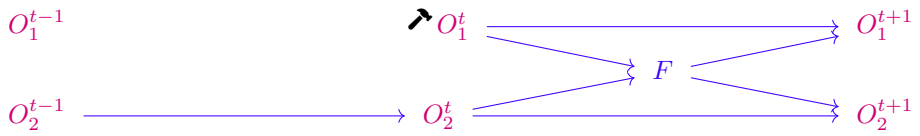
Yet, this representation still has drawbacks. Specifically, we break the identity of the force F into F_2 and F_3 which, in principle, can act as separate causal links (F_1 and F_4 can be thought of as inertia). This becomes apparent when interventions are applied. Let us imagine that somebody picks up object O_1 just before it collides with O_2. This can be represented by a do-calculus-like intervention applied to either O_1^t or O_1^{t+1}:

When the intervention is applied to O_1^t, the graph structure is preserved, thus implying a no-collision scenario (one of the balls was lifted). Yet, the same graph can also imply a collision scenario. This kind of setup necessitates having causal links between objects that can potentially collide irrespective of the actualization of said collision. While this approach can work in principle, it results in extremely dense graphs with complete subgraphs per time step, especially in cluttered scenes. Ideally, we would like to have causal links in our graph if there is an actualized interaction between the involved objects.

On the other hand, intervening on O_1^{t+1} results in a graph a with counter-intuitive interpretation: O_1 gets lifted at time step $t+1$, while O_2 behaves as if a collision has happened. This is due to the split of F into F_2 and F_3 since an intervention removes F_3 while leaving F_2 untouched. To tackle the aforementioned issues, let us re-imagine force edges as nodes and re-introduce F_2 and F_3 as a single node F and extend the time horizon by a step.



Now, imagine, just like before, the ball O_1 gets picked up at time step t. In do-calculus terms, this amounts to intervention to O_1^t which results in mutilation of the edge $O_1^{t-1} \to O_1^t$



While the problem of splitting of the force identity seems to be resolved here, the graph structure modeling the collision remains preserved despite the intervention. As mentioned before, this can be addressed by complete subgraphs per time step, which is not desirable for our purposes. This problem arises due to the inclusion of time dynamics into our graphs. Unlike in Pearlian Causality, in physics, interventions at a time step have implications for the causal connections corresponding to downstream time steps. To account for that, we have to re-imagine interventions under a new framework that takes physical processes and time into account (see Algorithm S1).

C Model details

C.1 Inductive bias

We introduce two inductive biases: (1) limiting each force node to interact with exactly two objects to reflect pairwise interactions, and (2) enforcing a bidirectional mirroring constraint to ensure temporal coherence in causal attribution. Formally the latter is defined as:

$$\begin{aligned}\forall i,j,k,t: \left\{E(O_i^t, F_j^{t+1}), E(O_k^t, F_j^{t+1})\right\} \subset G^{\mathcal{E}}_{\mathcal{O}^t:\mathcal{O}^{t+1}} \implies \\ E(F_j^{t+1}, O_i^{t+1}) \in G^{\mathcal{E}}_{\mathcal{O}^t:\mathcal{O}^{t+1}} \veebar E(F_j^{t+1}, O_k^{t+1}) \in G^{\mathcal{E}}_{\mathcal{O}^t:\mathcal{O}^{t+1}}, \\ \forall i,j,t: E(F_j^{t+1}, O_i^{t+1}) \in G^{\mathcal{E}}_{\mathcal{O}^t:\mathcal{O}^{t+1}} \implies E(O_i^t, F_j^{t+1}) \in G^{\mathcal{E}}_{\mathcal{O}^t:\mathcal{O}^{t+1}}.\end{aligned}$$

C.2 Vector constraints

Mutability is coded by $\overset{M}{\bullet}$. If $\overset{M}{\circ}$, the corresponding sub-vector does not change over time, i.e., an immutable sub-vector does not change over time: $\forall t, j : \left[\overset{M}{\circ}\right]^t_{F_j} = \left[\overset{M}{\circ}\right]^{t+1}_{F_j}$.

Causal relevance is coded by $\overset{C}{\bullet}$. Perturbing the sub-vectors of the parent with $\overset{C}{\circ}$ does not affect the child nodes, i.e., only the causally-relevant $\overset{C}{\bullet}$ sub-vectors affect the child nodes:

$$\forall t, i : \left[\overset{C}{\bullet}\right]^{t+1}_{F_j} = \left[\overset{C}{\bullet}\right]^{t+1}_{\widetilde{F}_j} \implies f_O\left(O_i^t, \left\{F_j^{t+1}\right\}_{j|(j,i)\in I^t}; \theta_O\right) = f_O\left(O_i^t, \left\{\widetilde{F}_j^{t+1}\right\}_{j|(j,i)\in I^t}; \theta_O\right).$$

Lastly, *control* relevance is coded by $\overset{K}{\bullet}$. Two force vectors whose sub-vectors with $\overset{K}{\bullet}$ are identical have identical control functions that are conditioned on them, i.e., control functions are conditioned only on the control-relevant sub-vectors :

$$\forall t : \left[\overset{K}{\bullet}\right]^{t+1}_{F_j} = \left[\overset{K}{\bullet}\right]^{t+1}_{\widetilde{F}_j} \implies \rho^t_{\mathcal{O}\leftrightarrow\widetilde{\mathcal{F}}} = \rho^t_{\mathcal{O}\leftrightarrow\mathcal{F}}.$$

C.3 Causal Process Block

Given the data $F^t, O_1^t, \ldots, O_n^t$, and the chosen indices J^t, we calculate F^{t+1} in the following way:

$$F^{t+1} := f_F\left(F^t, O_1^t, \ldots, O_n^t, J^t \mid \theta_F\right),$$

with O^{t+1} also calculated similarly.

$$
\begin{aligned}
\text{gate} &:= \frac{1}{|J^t|} \sum_{i \in J^t} \left(\left[\overset{C}{\bullet} \right]^t_{O_i} W^{\mathcal{F}}_{\text{gate}} \right) W^{\mathcal{F}}_{\text{output}}, \\
\text{residual} &:= \text{gate} + \left[\overset{M}{\bullet} \right]^t_F, \\
\left[\overset{M}{\bullet} \right]^{t+1}_F &:= \chi_{\text{gate}\neq 0} \odot \text{FFN}\left(\text{Norm}\left(\text{residual}\right)\right) + \text{residual}, \\
F^{t+1} &:= \left[\left[\overset{M}{\bullet} \right]^{t+1}_F ; \left[\overset{M}{\circ} \right]^t_F \right],
\end{aligned}
$$

where $\theta_F := \left\{ W^{\mathcal{F}}_{\text{gate}}, W^{\mathcal{F}}_{\text{output}}, W_1, W_2, b_1, b_2 \right\}$ FFN is a feed-forward neural network $\text{FFN}(x) := \max\left(0, xW_1 + b_1\right) W_2 + b_2$, $W^{\mathcal{F}}_{\text{gate}}$ is the analogue of the attention mechanisms value token projection, and $W^{\mathcal{F}}_{\text{output}}$ is again the analogous out-projection that maps the token from the attention dimension back to residual dimension (Vaswani et al., 2017).

D Additional Figures

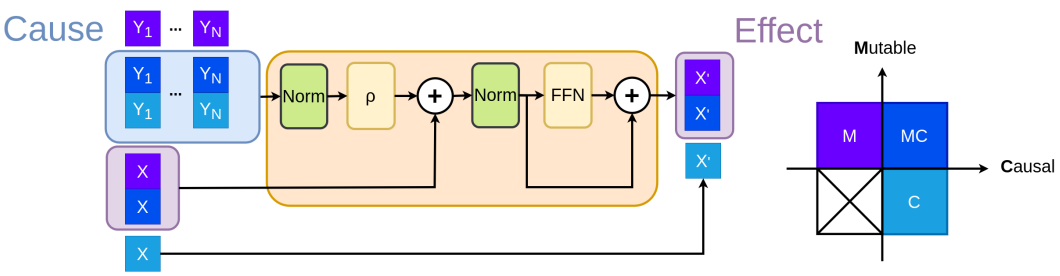


Figure S1: Causal Process Block.

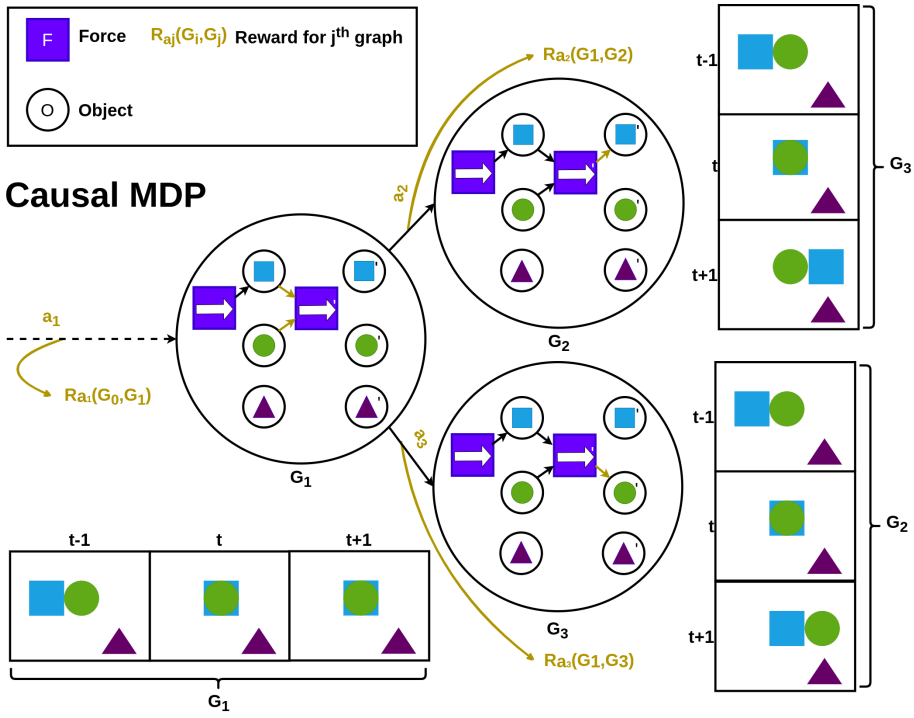


Figure S2: Causal MDP used by the reactive agents to construct causal process graphs.

E Plots

The y-axis in Figures $S6 - S8$ use Mean Reciprocal Rank, i.e., average inverse rank:

$$\text{MRR} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\text{rank}_n},$$

where rank_n is the rank of n^{th} sample over all reference state representations.

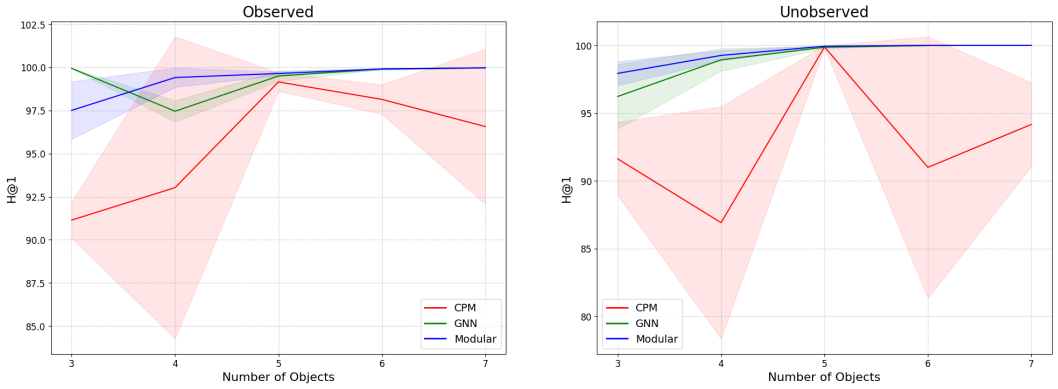


Figure S3: Prediction Metrics: Hit-at-One, 1-step.

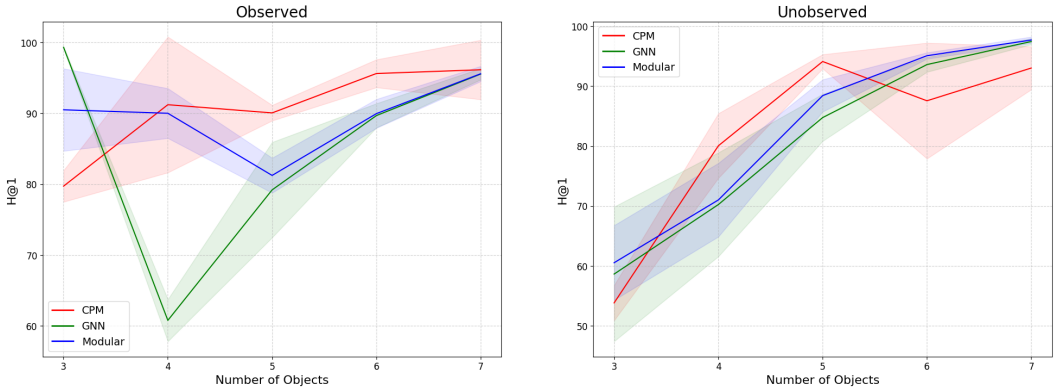


Figure S4: Prediction Metrics: Hit-at-One, 5-step.

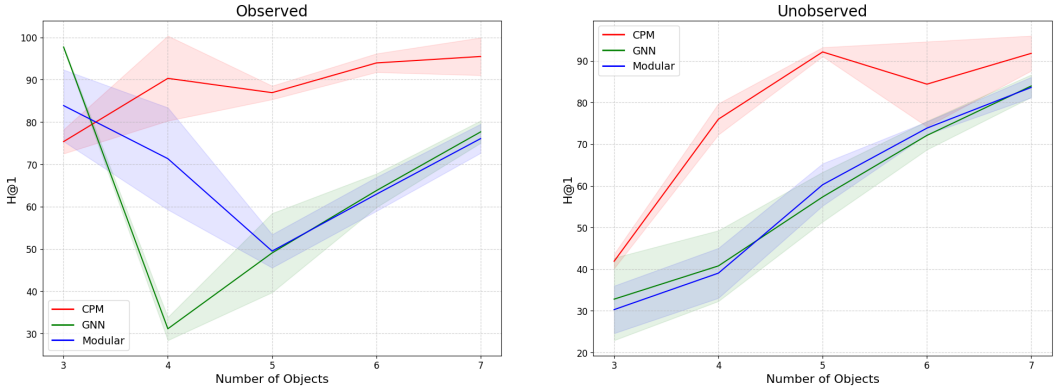


Figure S5: Prediction Metrics: Hit-at-One, 10-step.

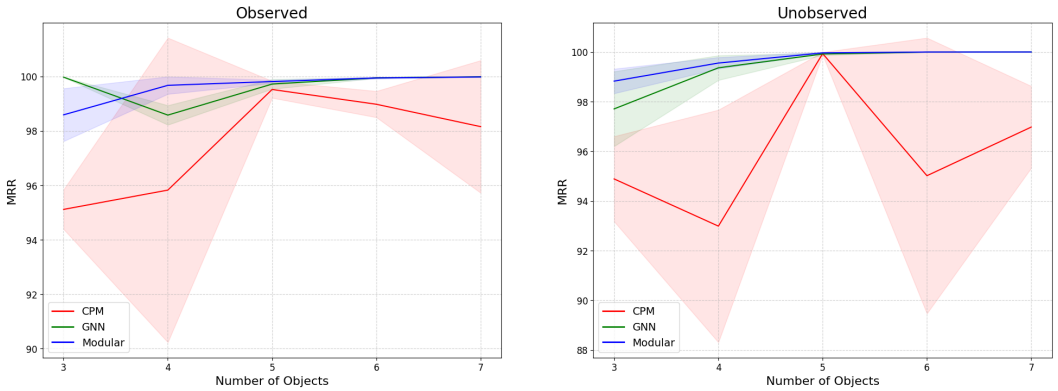


Figure S6: Prediction Metrics: Mean Reciprocal Rank, 1-step.

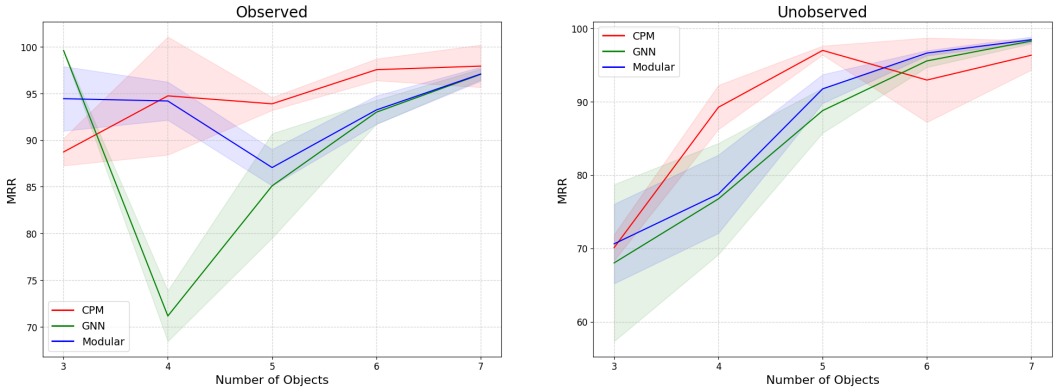


Figure S7: Prediction Metrics: Mean Reciprocal Rank, 5-step.

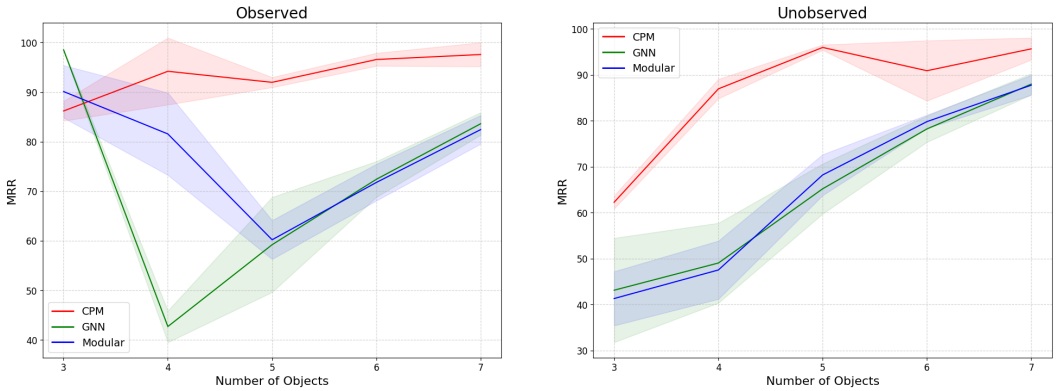


Figure S8: Prediction Metrics: Mean Reciprocal Rank, 10-step.

F Algorithms

Algorithm S1 Interventions under Causal Process Framework

Require: $\mathcal{F}^t, \mathcal{O}^t, f_F, f_O, \rho^t_{\mathcal{O}\leftrightarrow\mathcal{F}}, \rho^t_{\mathcal{O}}, act\left(F^{\tilde{t}}_*\right), G_{\mathcal{O}^{1:}\mathcal{O}^T}, t \in \{1,\dots,T\}, i \in \{1,\dots,I\}, j \in \{1,\dots,J\}$

Ensure: Output result $\widetilde{G}_{\mathcal{O}^{1:}\mathcal{O}^T}$

Initialize $\widetilde{\mathcal{F}}^{\tilde{t}} := \left\{F^{\tilde{t}}_*\right\} \dot{\cup} \mathcal{F}^{\tilde{t}}$

Initialize $\widetilde{\mathcal{O}}^{\tilde{t}-1} := \mathcal{O}^{\tilde{t}-1}$

Initialize $J^t := \{(i,j)\}$ s.t. $E\left(O^{\tilde{t}-1}_i, F^{\tilde{t}}_j\right) \in G^{\mathcal{E}}_{\mathcal{O}^{\tilde{t}-1}:\mathcal{F}^{\tilde{t}}}$

Initialize $\widetilde{G}_{\mathcal{O}^{1:}\mathcal{O}^{\tilde{t}-1}} := G_{\mathcal{O}^{1:}\mathcal{O}^{\tilde{t}-1}}$

for $t = \tilde{t},\dots,T$ **do** ▷ Loop from $t = \tilde{t}$ up to T

  $\widetilde{G}_{\mathcal{O}^{1:}\mathcal{F}^t} := \left(\widetilde{\mathcal{F}}^t \dot{\cup} \widetilde{G}^{\mathcal{V}}_{\mathcal{O}^{1:}\mathcal{O}^{t-1}}, \left\{E\left(\widetilde{O}^{t-1}_i, \widetilde{F}^t_j\right)\right\}_{(i,j)\in J^t} \dot{\cup} \widetilde{G}^{\mathcal{E}}_{\mathcal{O}^{1:}\mathcal{O}^{t-1}}\right)$ ▷ Update the graph

  $I^t \sim \rho^t_{\widetilde{\mathcal{O}}\leftrightarrow\widetilde{\mathcal{F}}}$ ▷ Sample new edges

  for $i = 1,\dots,I$ **do** ▷ Loop from $i = 1$ up to I

    $\widetilde{O}^t_i := f_O\left(\widetilde{O}^{t-1}_i, \left\{\widetilde{F}^t_j\right\}_{j|(j,i)\in I^t}\right)$ ▷ Update the nodes

  end for

  $\widetilde{G}_{\mathcal{O}^{1:}\mathcal{O}^t} := \left(\widetilde{\mathcal{O}}^t \dot{\cup} \widetilde{G}^{\mathcal{V}}_{\mathcal{O}^{1:}\mathcal{F}^t}, \left\{E\left(\widetilde{F}^t_j, \widetilde{O}^t_i\right)\right\}_{(j,i)\in I^t} \dot{\cup} \widetilde{G}^{\mathcal{E}}_{\mathcal{O}^{1:}\mathcal{F}^t}\right)$ ▷ Update the graph

  $J^t \sim \rho^t_{\widetilde{\mathcal{O}}}$ ▷ Sample new edges

  for $j = 1,\dots,J$ **do** ▷ Loop from $j = 1$ up to J

    $\widetilde{F}^{t+1}_j := f_F\left(\widetilde{F}^t_j, \left\{\widetilde{O}^t_i\right\}_{i|(i,j)\in J^t}\right)$ ▷ Update the nodes

  end for

end for

return $\widetilde{G}_{\mathcal{O}^{1:}\mathcal{O}^T}$

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Turan Orujlu.

References

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=i80OPhOCVH2.

Federico Barbero, Andrea Banino, Steven Kapturowski, Dharshan Kumaran, João G. M. Araújo, Alex Vitvitskyi, Razvan Pascanu, and Petar Velickovic. Transformers need glasses! information over-squashing in language tasks. *CoRR*, abs/2406.04267, 2024a. DOI: 10.48550/ARXIV.2406.04267. URL https://doi.org/10.48550/arXiv.2406.04267.

Federico Barbero, Ameya Velingker, Amin Saberi, Michael M. Bronstein, and Francesco Di Giovanni. Locality-aware graph rewiring in gnns. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024b. URL https://openreview.net/forum?id=4Ua4hKiAJX.

Elias Bareinboim, S Lee, and J Zhang. An introduction to causal reinforcement learning, 2021.

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinícius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Çaglar Gülçehre, H. Francis Song, Andrew J. Ballard, Justin Gilmer, George E. Dahl, Ashish Vaswani, Kelsey R. Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matthew M. Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. *CoRR*, abs/1806.01261, 2018. URL http://arxiv.org/abs/1806.01261.

Lars Buesing, Theophane Weber, Yori Zwols, Nicolas Heess, Sébastien Racanière, Arthur Guez, and Jean-Baptiste Lespiau. Woulda, coulda, shoulda: Counterfactually-guided policy search. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL https://openreview.net/forum?id=BJG0voC9YQ.

Martin V. Butz, Maximilian Mittenbühler, Sarah Schwöbel, Asya Achimova, Christian Gumbsch, Sebastian Otte, and Stefan Kiebel. Contextualizing predictive minds. *Neuroscience Biobehavioral Reviews*, 168:105948, 2025. ISSN 0149-7634. DOI: https://doi.org/10.1016/j.neubiorev.2024.105948. URL https://www.sciencedirect.com/science/article/pii/S0149763424004172.

Hanna M Dettki, Brenden M Lake, Charley M Wu, and Bob Rehder. Do large language models reason causally like us? even better? *arXiv preprint arXiv:2502.10215*, 2025.

Phil Dowe. *Physical Causation*. New York, 2000.

Maxime Gasse, Damien Grasset, Guillaume Gaudron, and Pierre-Yves Oudeyer. Using confounded data in latent model-based reinforcement learning. *Trans. Mach. Learn. Res.*, 2023, 2023. URL https://openreview.net/forum?id=nFWRuJXPkU.

Tobias Gerstenberg, Noah D. Goodman, David A. Lagnado, and Joshua B. Tenenbaum. A counterfactual simulation model of causal judgments for physical events. *Psychological review*, 2020. URL https://api.semanticscholar.org/CorpusID:235361720.

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Lio, and Michael M. Bronstein. On over-squashing in message passing neural networks: The impact of width, depth, and topology. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 7865–7885. PMLR, 2023. URL https://proceedings.mlr.press/v202/di-giovanni23a.html.

Francesco Di Giovanni, T. Konstantin Rusch, Michael M. Bronstein, Andreea Deac, Marc Lackenby, Siddhartha Mishra, and Petar Velickovic. How does over-squashing affect the power of gnns? *Trans. Mach. Learn. Res.*, 2024, 2024. URL https://openreview.net/forum?id=KJRoQvRWNs.

Christian Gumbsch, Martin V Butz, and Georg Martius. Sparsely changing latent states for prediction and planning in partially observable domains. *Advances in Neural Information Processing Systems*, 34:17518–17531, 2021.

Nan Rosemary Ke, Aniket Didolkar, Sarthak Mittal, Anirudh Goyal, Guillaume Lajoie, Stefan Bauer, Danilo Jimenez Rezende, Michael Mozer, Yoshua Bengio, and Chris Pal. Systematic evaluation of causal discovery in visual model based reinforcement learning. In Joaquin Vanschoren and Sai-Kit Yeung (eds.), *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021, virtual*, 2021. URL https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/8f121ce07d74717e0b1f21d122e04521-Abstract-round2.html.

Thomas N. Kipf, Elise van der Pol, and Max Welling. Contrastive learning of structured world models. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=H1gax6VtDB.

Richard D. Lange and Konrad P. Kording. Causality in the human niche: lessons for machine learning. *CoRR*, abs/2506.13803, 2025. DOI: 10.48550/ARXIV.2506.13803. URL https://doi.org/10.48550/arXiv.2506.13803.

Anson Lei, Bernhard Schölkopf, and Ingmar Posner. Spartan: A sparse transformer learning local causation. *arXiv preprint arXiv:2411.06890*, 2024.

Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Causal transformer for estimating counterfactual outcomes. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 15293–15329. PMLR, 2022. URL https://proceedings.mlr.press/v162/melnychuk22a.html.

Eshaan Nichani, Alex Damian, and Jason D. Lee. How transformers learn causal structure with gradient descent. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=jNM4imlHZv.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009. ISBN 052189560X.

Silviu Pitis, Elliot Creager, and Animesh Garg. Counterfactual data augmentation using locally factored dynamics. *Advances in Neural Information Processing Systems*, 33:3976–3990, 2020.

James Robins and Miguel Hernan. *Estimation of the causal effects of time-varying exposure*, pp. 553–599. 08 2008. ISBN 978-1-58488-658-7. DOI: 10.1201/9781420011579.ch23.

Raanan Y. Rohekar, Yaniv Gurwicz, and Shami Nisimov. Causal interpretation of self-attention in pre-trained transformers. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/642a321fba8a0f03765318e629cb93ea-Abstract-Conference.html.

Donald B. Rubin. Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, 6(1):34 – 58, 1978. DOI: 10.1214/aos/1176344064. URL https://doi.org/10.1214/aos/1176344064.

Bertrand Russell. *Human Knowledge: Its Scope and Limits*. Routledge, London and New York, 1948.

Wesley C. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, 1984.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Trans. Neural Networks*, 20(1):61–80, 2009. DOI: 10.1109/TNN.2008.2005605. URL https://doi.org/10.1109/TNN.2008.2005605.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Maximilian Seitzer, Bernhard Schölkopf, and Georg Martius. Causal influence detection for improving efficiency in reinforcement learning. *Advances in Neural Information Processing Systems*, 34: 22905–22918, 2021.

Xiao Shou, Debarun Bhattacharjya, Tian Gao, Dharmashankar Subramanian, Oktie Hassanzadeh, and Kristin P. Bennett. Pairwise causality guided transformers for event sequences. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/91b047c5f5bd41ef56bfaf4ad0bd19e3-Abstract-Conference.html.

Brian Skyrms. Causal necessity. *Philosophy of Science*, 48(2):329–335, 1981. DOI: 10.1086/289003.

Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=7UmjRGzp-A.

Núria Armengol Urpí, Marco Bagatella, Marin Vlastelica, and Georg Martius. Causal action influence aware counterfactual data augmentation. *arXiv preprint arXiv:2405.18917*, 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 5998–6008, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Christopher J. C. H. Watkins and Peter Dayan. Technical note q-learning. *Mach. Learn.*, 8:279–292, 1992. DOI: 10.1007/BF00992698. URL https://doi.org/10.1007/BF00992698.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach. Learn.*, 8:229–256, 1992. DOI: 10.1007/BF00992696. URL https://doi.org/10.1007/BF00992696.

Moritz Willig, Tim Nelson Tobiasch, Florian Peter Busch, Jonas Seng, Devendra Singh Dhami, and Kristian Kersting. Systems with switching causal relations: A meta-causal perspective. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL https://openreview.net/forum?id=J9VogDTa1W.

Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection: Expressiveness, learnability, and inference. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 10823–10836, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/5989add1703e4b0480f75e2390739f34-Abstract.html.

Matej Zecevic, Devendra Singh Dhami, Petar Velickovic, and Kristian Kersting. Relating graph neural networks to structural causal models. *CoRR*, abs/2109.04173, 2021. URL https://arxiv.org/abs/2109.04173.