Learning Nonlinear Causal Reductions to Explain RL Policies

ICLR 2026

Poster Session 2: P3-#106

Armin Kekić, Jan Schneider, Dieter Büchler,

Bernhard Schölkopf*, Michel Besserve*

*Joint supervision

23rd April 2026

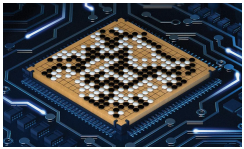
MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

Reinforcement Learning

1 Motivation

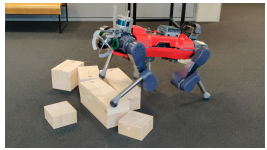
Games

Mnih et al., “Human-level control through deep reinforcement learning”, **Nature** (2015).

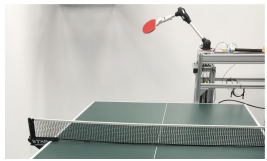


Silver et al., “Mastering the game of Go without human knowledge”, **Nature** (2017).

Robotics

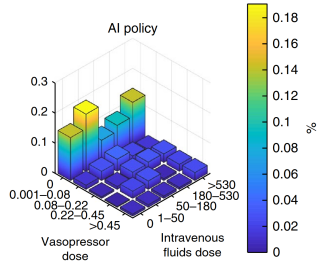


Rudin et al., “Learning to walk in minutes using massively parallel deep reinforcement learning”, **CoRL** (2022).



Büchler et al., “Learning to play table tennis from scratch using muscular robots”, **T-Ro** (2022).

Healthcare



Komorowski et al., “The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care”, **Nature Medicine** (2018).

Reinforcement Learning

1 Motivation

- **Understanding the behavior of RL policies** important for reliability, safety, and trust
- Policies have many parameters, state and action spaces can be high-dimensional $\rightarrow$ **difficult to understand**
- **Goal:** find high-level explanations for why a policy fails or succeeds

Causal Model Reduction is a principled way to learn high-level causes for effects in complex systems

Main Idea

1 Motivation

We treat RL rollouts as a **low-level causal model** and **learn a map** to a smaller **high-level causal representation**.

Causal Models

2 Background

Structural Causal Model (SCM)

An n-dim. SCM is a triplet $\mathcal{M} = (\mathcal{G}, \mathbb{S}, P_{\mathbf{U}})$ consisting of

- a directed graph $\mathcal{G}$ with n vertices;
- a set of structural equations $\mathbb{S} = \{X_j := f_j(\mathbf{PA}_j^{\mathcal{G}}, U_j), \quad j = 1, \ldots, n\}$;
- a distribution $P_{\mathbf{U}}$ over exogenous variables $\{U_j\}_{j \leq n}$.

The SCM entails a joint distribution $P_{\mathbf{X}}$ over endogenous variables.

An **intervention $\mathbf{i}$** changes some/all structural equations in $\mathbb{S}$, inducing $P_{\mathbf{X}}^{\mathbf{i}}$.

Maps between Causal Models

Causal Consistency[1] (informal)

Let

- $\mathcal{L}$ be an SCM with variables $\mathbf{X} \in \mathcal{X}$ and intervention set $\mathcal{I}$;
- $\mathcal{H}$ be an SCM with variables $\mathbf{Z} \in \mathcal{Z}$ and intervention set $\mathcal{J}$;
- $\tau : \mathcal{X} \to \mathcal{Z}$ surjective.

If $\exists\, \omega : \mathcal{I} \to \mathcal{J}$ s.t.

$$P_{\mathcal{H}}^{\omega(i)}(\tau(\mathbf{X})) = \tau_{\#}(P_{\mathcal{L}}^{i}(\mathbf{X}))\,,$$

we call τ **causally consistent**.

$$\begin{array}{ccc} P_{\mathcal{L}}(\mathbf{X}) & \xrightarrow{\ \tau\ } & P_{\mathcal{H}}(\tau(\mathbf{X})) \\ \Big\downarrow i & & \Big\downarrow \omega(i) \\ P_{\mathcal{L}}^{i}(\mathbf{X}) & \xrightarrow[\ \tau\ ]{} & P_{\mathcal{H}}^{\omega(i)}(\tau(\mathbf{X})) \end{array}$$

[1] Rubenstein et al., "Causal Consistency of Structural Equation Models", **UAI** (2017).

Maps between Causal Models

Approximate Causal Consistency[2] (informal)

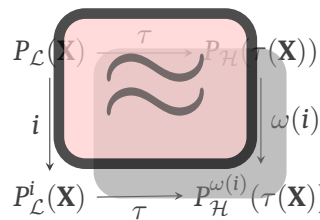
Let

- $\mathcal{L}$ be an SCM with variables $\mathbf{X} \in \mathcal{X}$ and intervention set $\mathcal{I}$;
- $\mathcal{H}$ be an SCM with variables $\mathbf{Z} \in \mathcal{Z}$ and intervention set $\mathcal{J}$;
- $\tau : \mathcal{X} \to \mathcal{Z}$ surjective.

If $\exists\, \omega : \mathcal{I} \to \mathcal{J}$ s.t.

$$P_{\mathcal{H}}^{\omega(i)}(\tau(\mathbf{X})) \approx \tau_{\#}(P_{\mathcal{L}}^{i}(\mathbf{X}))\,,$$

we call τ **approximately causally consistent**.

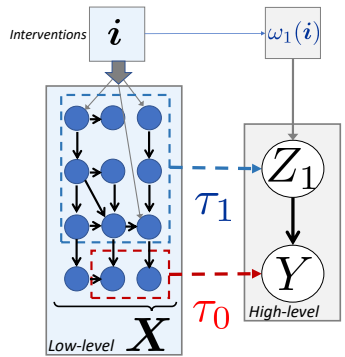


[2] Beckers et al., "Approximate causal abstractions", **UAI** (2020).

Linear Targeted Causal Reduction (TCR)[3]

- High-level variables:
  - scalar target $Y = \tau_0(\mathbf{X})$ (τ_0 fixed and known),
  - high-level cause $Z_1 = \tau_1(\mathbf{X})$ (τ_1 learned, linear)
- Shift interventions:
  $X_k := f_k(\mathbf{PA}_k, U_k) \;\mapsto\; X_k := f_k(\mathbf{PA}_k, U_k) + i_k$
- Linear intervention map $\omega_1(\mathbf{i})$
- Parametrized high-level SCM: linear Gaussian
- Causal Consistency Loss:

$$\mathcal{L}_{\text{cons}} = \mathbb{E}_{\mathbf{i}\sim P(\mathbf{i})}\left[\mathrm{KL}\Big(\underbrace{\widehat{P}_{\tau}^{(\mathbf{i})}(Y,Z)}_{\substack{\text{pushforward distr.}\\ \text{through } \tau}} \;\Big\|\; \underbrace{P_{\mathcal{H}}^{(\omega(\mathbf{i}))}(Y,Z)}_{\substack{\text{high-level}\\ \text{interventional distr.}}} \Big)\right]$$



[3] Kekić et al., "Targeted Reduction of Causal Models", **UAI** (2024).

From RL to Causal Model Reduction

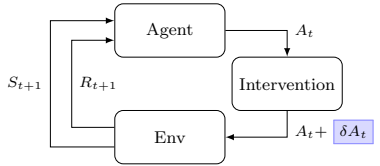
3 TCR for RL

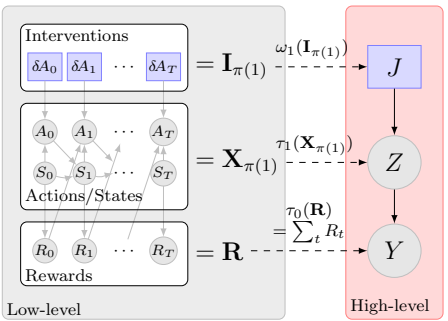
Low-level causal model: State-, action- and reward-variables

Shift interventions: Actions A_t selected by policy are perturbed by small random shift δA_t

Target: Cumulative rewards

→ Learn high-level SCM that explains changes in the cumulative reward





Nonlinear TCR (nTCR)

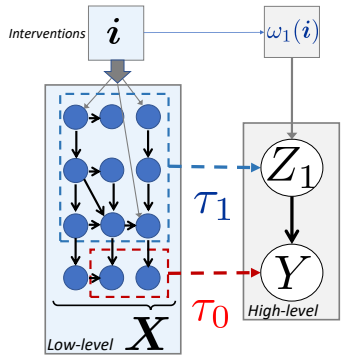
Linear τ and ω limit expressivity

$\rightarrow$ nTCR: Allow τ and ω to be **nonlinear**

$\rightarrow$ High-level SCM stays linear Gaussian for interpretability

Challenges:

- $\mathcal{L}_{\text{cons}}$ can lead to complicated non-Gaussian high-level distributions, even if low-level is Gaussian
- General nonlinear τ- and ω-transformations are difficult to interpret

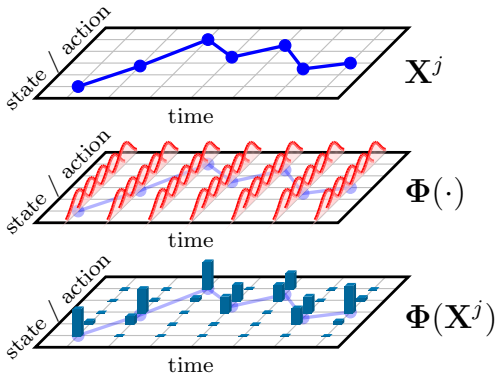


Interpretable Nonlinear Function Class

Nonlinear variable transformation with **Gaussian kernels**:

$$\tau_1(\mathbf{X}) = \sum_{j=1}^{d} \sum_{t=1}^{T} w_{j,t} \cdot \underbrace{\exp\left(-(x-\mu_{j,t})^2/2\sigma_{j,t}^2\right)}_{\phi_{j,t}(X_t^j)}$$

- Learned weights $w_{j,t}$ show which features at which times matter most
- Middle ground between linear maps and full nonlinearity



Pendulum Task[4]

State:
position, angular velocity

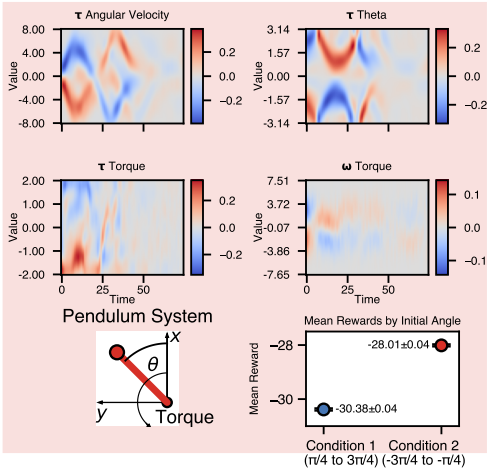
Action:
torque

[4] Towers et al., “Gymnasium: A Standard Interface for Reinforcement Learning Environments”, **arXiv** (2024).

Pendulum Task

4 Case Studies

Policy A



Policy B

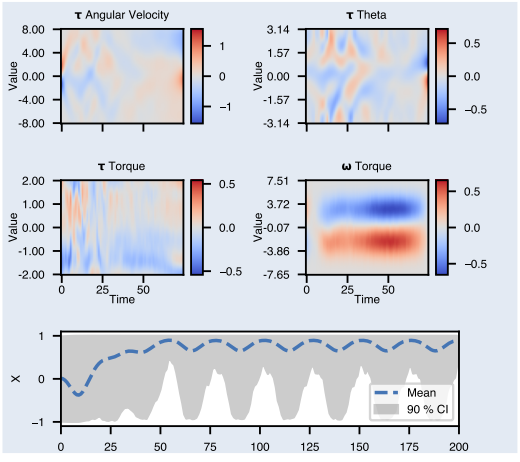
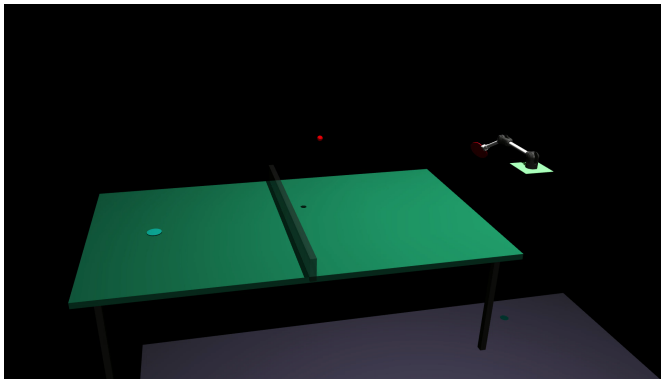


Table Tennis Task[5]

4 Case Studies



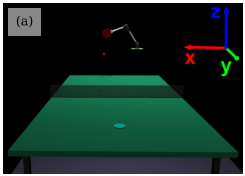
State:
- ball position, velocity
- joint positions, velocities,
- muscle pressures

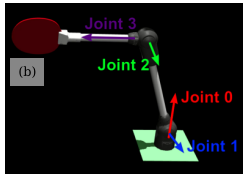
Action:
- pressure commands

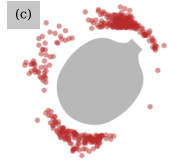
[5] Büchler et al., “Learning to play table tennis from scratch using muscular robots”, **T-Ro** (2022).

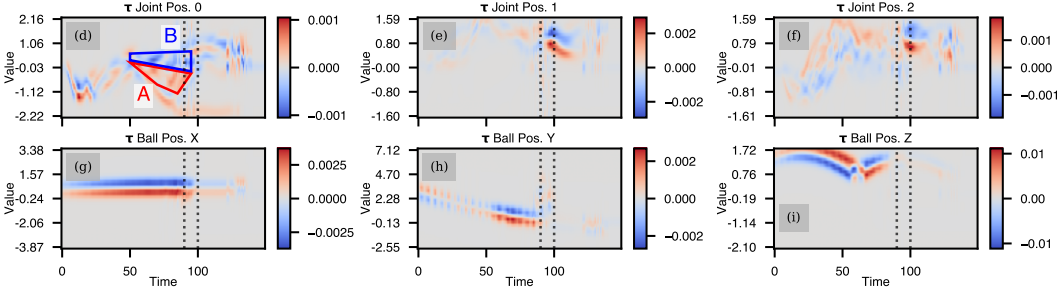
Table Tennis Task

4 Case Studies









Learning Nonlinear Causal Reductions to Explain RL Policies

5 Summary

- Treat policy rollouts as a **low-level causal model**
- Define cumulative reward as **target**
- Using **low-level interventions**, we learn mapping to a **high-level causal model** to explain causes of the target
- Visualizing the mapping uncovers biases and failure modes of the policy