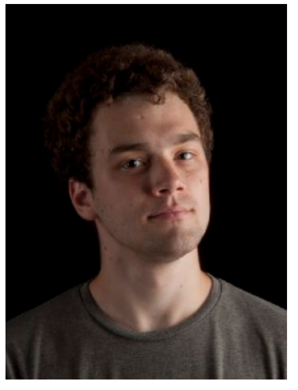
# Causal Imitation Learning with Unobserved Confounders



Junzhe Zhang, Daniel Kumor, Elias Bareinboim
Causal Artificial Intelligence Lab
Columbia University

Thirty-fourth Conference on Neural Information Processing Systems, 2020

# Motivation

- Imitation Learning: learning a policy from demonstrations of an expert so that it achieves the expert's performance.

- Challenge: reward signal is unobserved.

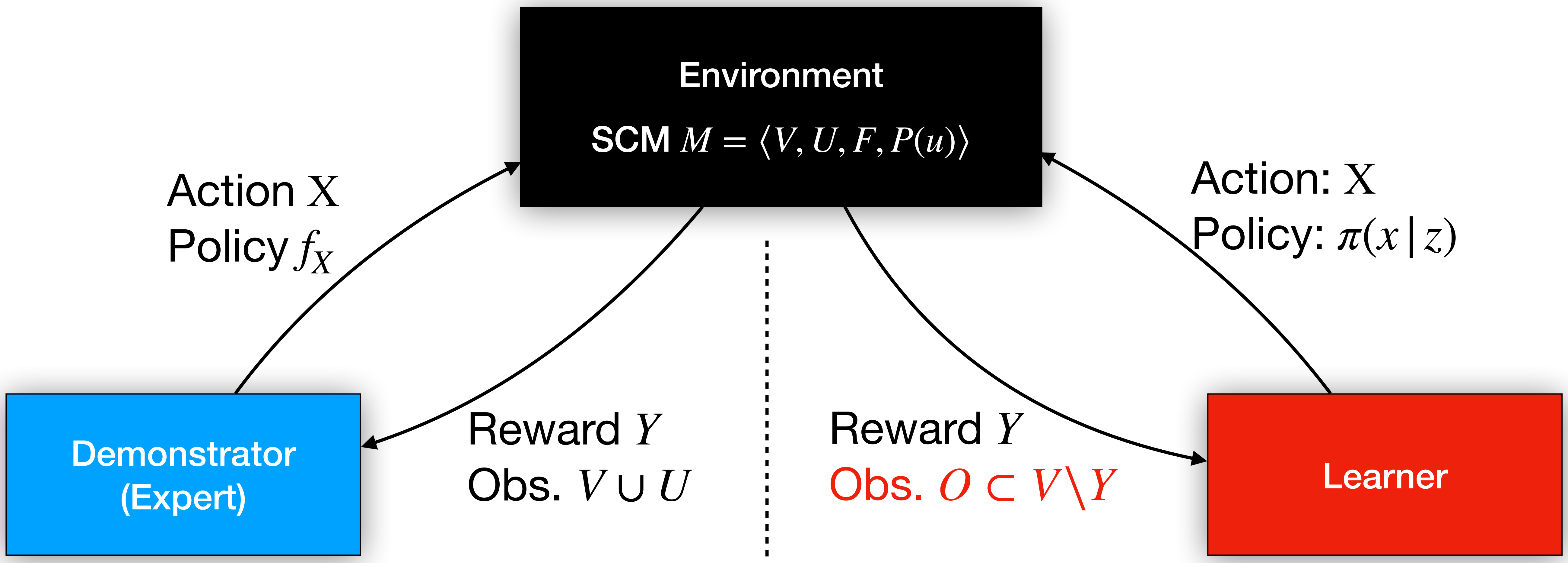
- Assumption: the expert and the learner share the *same state-action space*.

  1. Behavior Cloning

  2. Inverse Reinforcement Learning

- Goal: Perform imitation learning when some input variables of the demonstrator's policy are unobserved.

# Fundamental Problem of Causal Imitation Learning (FPIL)




Summary

Observational Data: $P(o)$

Performance of expert: $E[Y]$

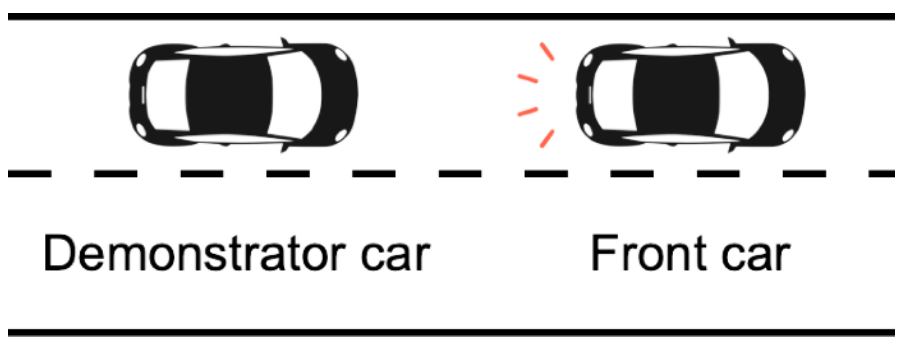
Input for learner: $P(o)$

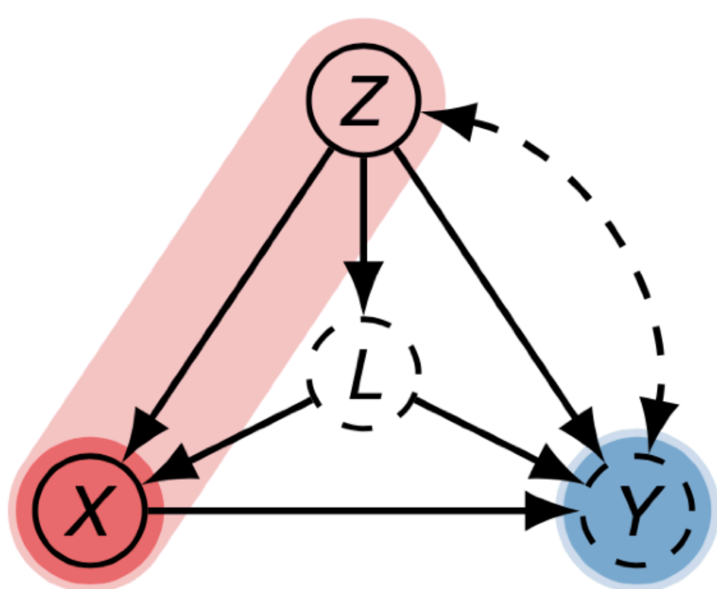
Task: find *an imitating policy* $\pi(x \mid z)$ s.t.

$$E_M[Y \mid do(\pi)] = E_M[Y]$$

given obs. $P(o)$ (not including $Y$)

# Imitation from Aerial Driving Footage

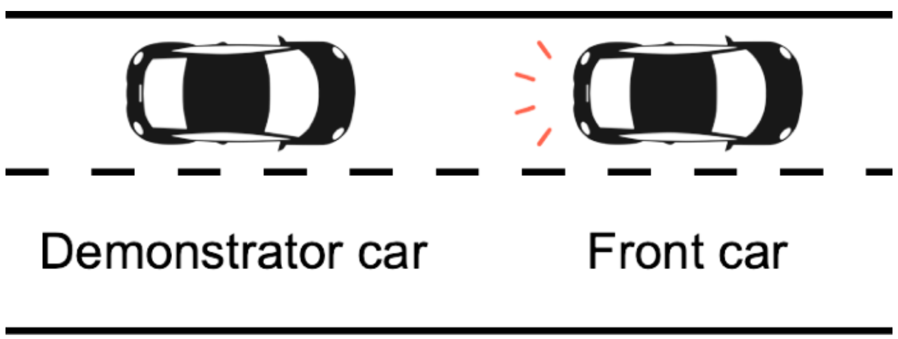


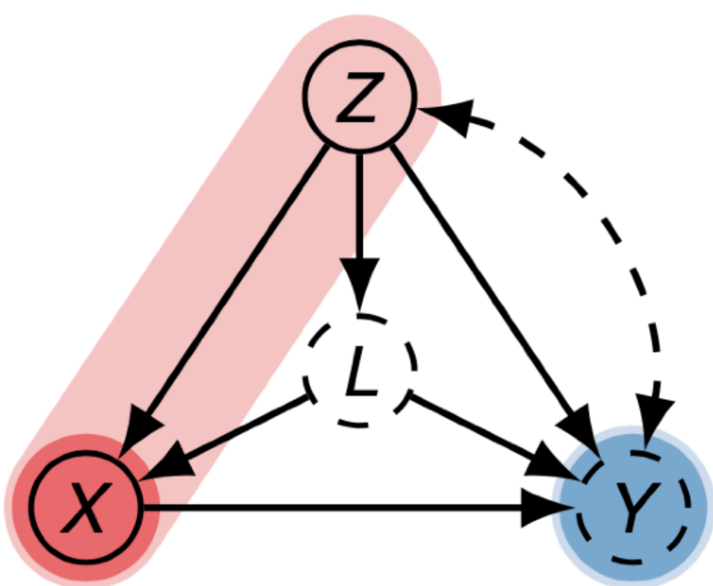




Causal Diagram $G$

- $X$: acceleration of the demonstrator car
- $Y$: driving performance (latent)
- $Z$: velocity and locations of both cars
- $L$: tail light of the front car (latent)
- $P(x, z)$: observational distribution
- $\Pi = \{\pi(x \mid z)\}$: learner's policy space

# Imitation from Aerial Driving Footage




Causal Diagram $G$

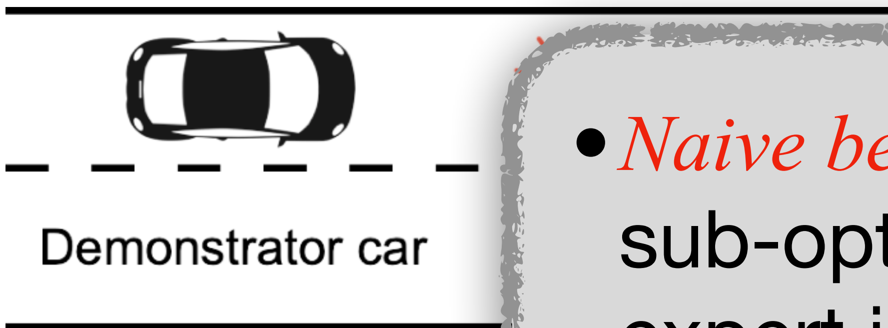
- Consider a SCM as follows

$$X \leftarrow \neg L \oplus Z;$$
$$Y \leftarrow X \oplus L \oplus Z.$$

$Z, L$ are drawn uniformly over $\{0,1\}$.

- $E[Y] = P(Y=1) = 1$
- Behavior cloning $\pi(x \mid z) = P(x \mid z)$
- $E[Y \mid do(\pi)] = P(Y = 1 \mid do(\pi)) = 0.5$

# Imitation from Aerial Driving Footage



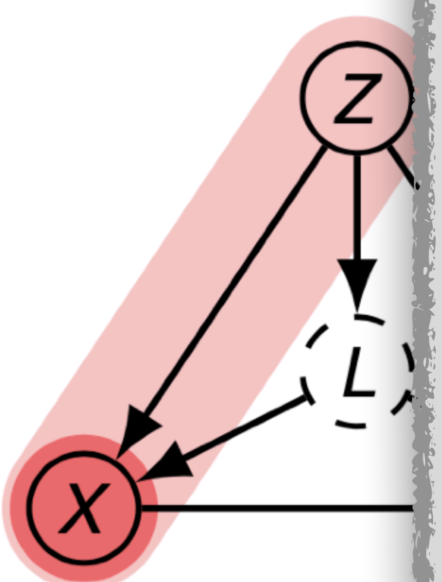

- Consider a SCM as follows

over $\{0,1\}$.

$= P(x \mid z)$

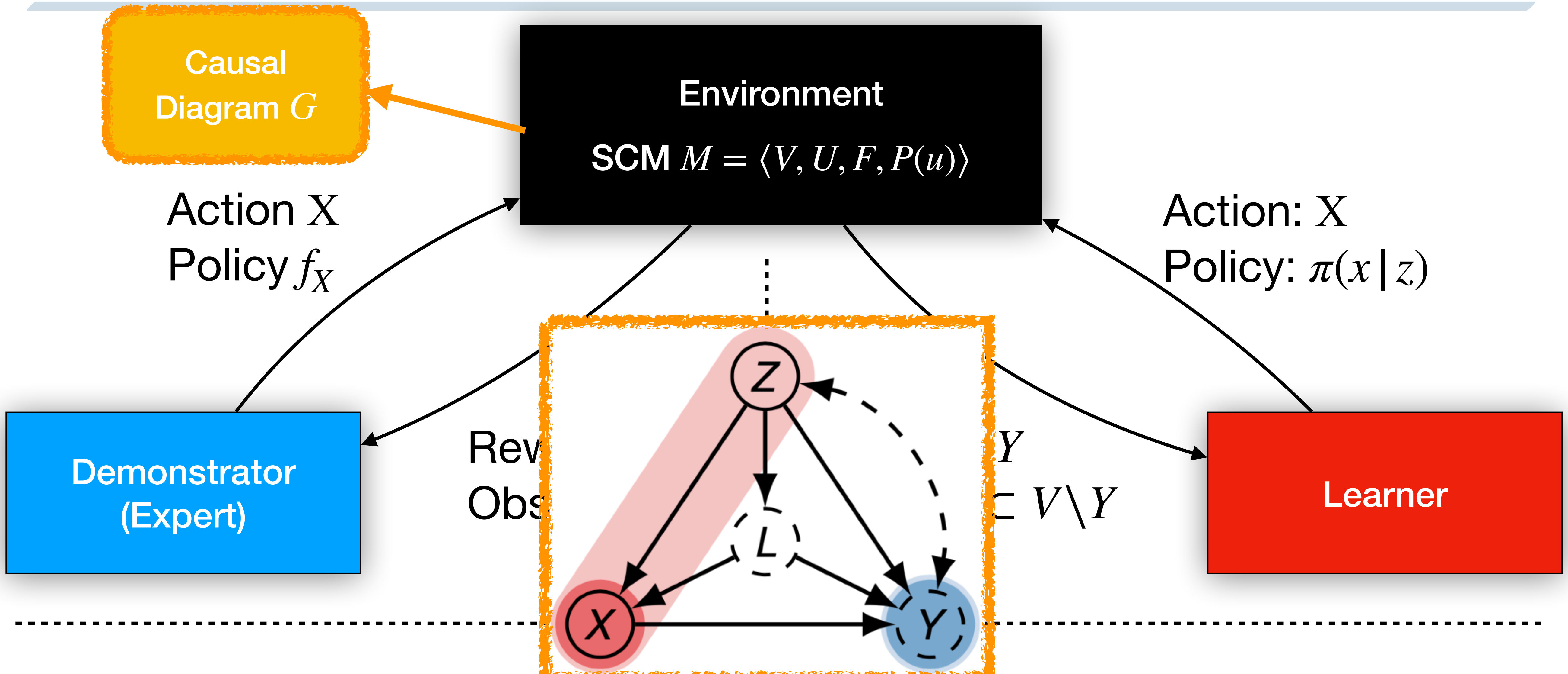
$\pi)) = 0.5$

- *Naive behavior cloning* could lead to sub-optimal policy even when the expert is optimal & cloning is perfect.
- Behavior cloning does not guarantees the successful imitation given the observational data alone.
- Solution: exploring causal relationships of the underlying environment.




Causal Diagram $G$

# Fundamental Problem of Causal Imitation Learning (FPIL)




Summary

Observational Data: $P(o)$

Performance of expert: $E[Y]$

Input for learner: $P(o)$ and $G$.

Task: find *an imitating policy* $\pi(x \mid z)$ s.t.

$$E_M[Y \mid do(\pi)] = E_M[Y]$$

given obs. $P(o)$ and diagram $G$.

# Big Picture / Contributions

- We model imitation through causal semantics, and note that behavior cloning does NOT always lead to successful imitation. The natural question is then whether, or when, imitation could work, in particular, behavior cloning.

- Question 1 (Sec. 2). Causal Behavioral Cloning (BC)

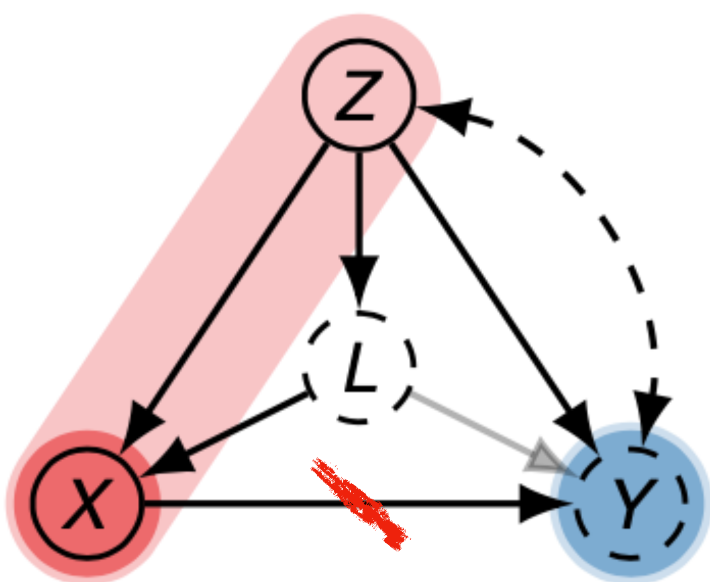
  - Under what conditions behavior cloning works?

- Question 2 (Sec. 3). Beyond Causal BC

  - We introduce a novel family of imitation methods that can lead to successful imitation even when BC is provably inefficient.

# Imitation by Feature Selection

- Theorem (Imitation by Backdoor): $P(y)$ is imitable if there exists a set of nodes $Z'$ satisfies *backdoor criterion*: $Z' \subseteq Z$ and $\left(X \perp Y \mid Z'\right)$ in a subgraph $G_{\underline{X}}$ with outgoing arrows of $X$ removed.

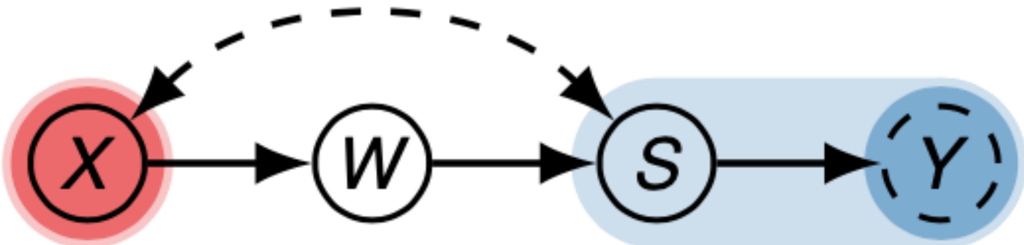



Causal Diagram $G$

Behavior Cloning

- Assume that tail light $L$ does not affect driving performance $Y$
- $Z$ is backdoor admissible
- $P(y)$ is imitable by adjustment on $Z$
- The imitating policy $\pi(x \mid z) = P(x \mid z)$

# Moving Beyond Behavior Cloning



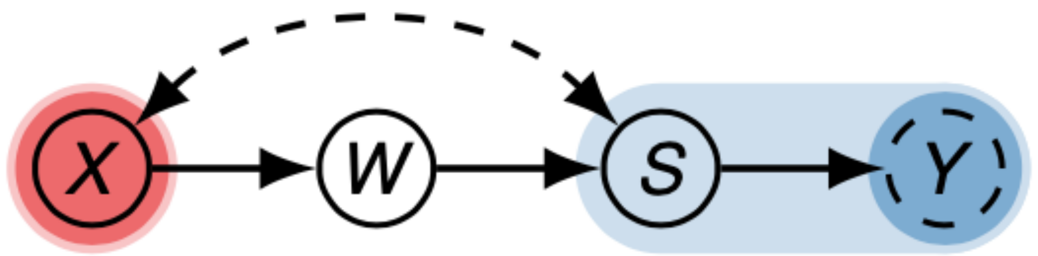

- Observational distribution $P(x, w, s)$
- Policy space $\Pi = \{\pi(x)\}$

- $S$ is a surrogate: $P(s \mid do(\pi)) = P(s) \Rightarrow P(y \mid do(\pi)) = P(y)$
- Imitating policy $\pi$: $P(s \mid do(\pi)) = \sum_x P(s \mid do(x))\pi(x) = P(s)$
- For binary $X, W, S$, $\pi(x_1) = \dfrac{P(s_1 \mid do(x_1)) - P(s_1)}{P(s_1 \mid do(x_1)) - P(s_1 \mid do(x_0))}$

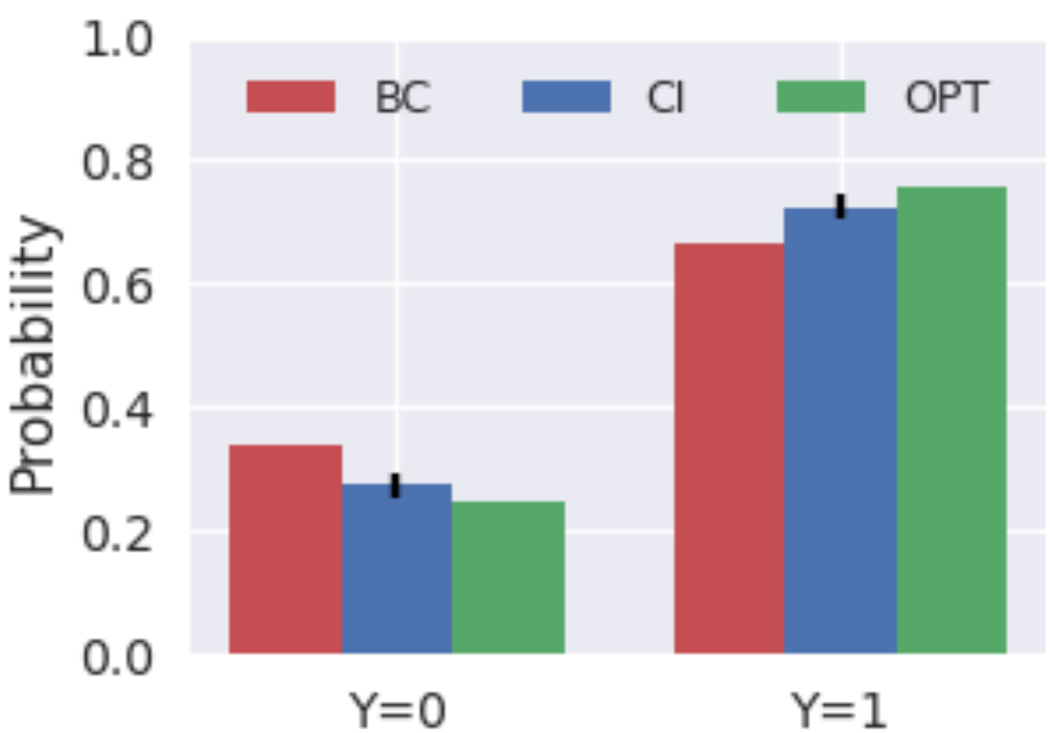
$$\boxed{P(s \mid do(x)) = \sum_w P(w \mid x) \sum_{x'} P(s \mid w, x')P(x')}$$

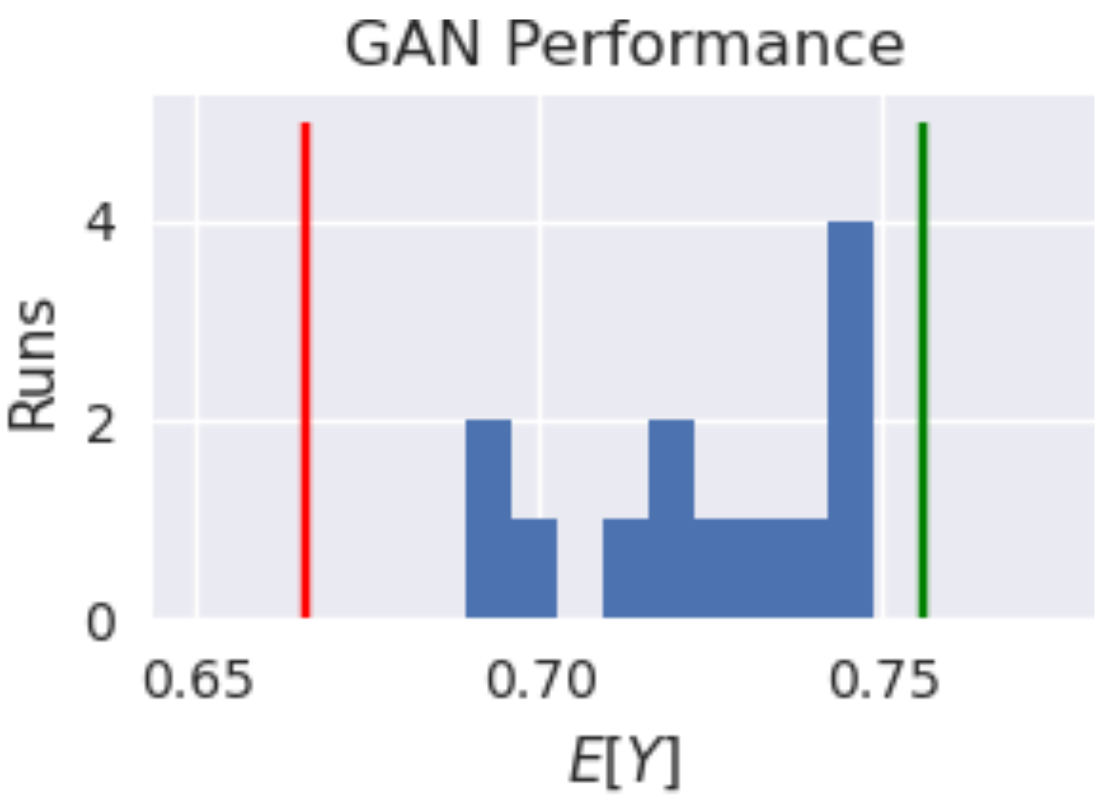
Causal Identification

# Simulations




- $X, S, Y$ are binary variables
- $W$ is a MNIST digits



# Conclusion

- Formulating imitation learning in the semantics of structural causal models.
- Conditions under which imitation learning is feasible.
- What is in the paper (contributions):
    - Complete algorithms for finding backdoor admissible sets for behavior cloning.
    - Algorithms for finding other instruments for imitation, beyond behavior cloning.
    - Optimization procedure based on GANs for solving for imitating policies in high-dimensional domains.