TECHNICAL REPORT
R-145
Mar, 2026

Sequential Causal Games

Aurghya Maiti [1] **Elias Bareinboim** [1]

Abstract

Sequential multi-agent decision-making is commonly modeled using extensive-form, repeated, or Markov games. While powerful, these frameworks fail to capture the causal relations among the system's variables and lack semantics for interventions, latent confounding, and counterfactual reasoning. We introduce Sequential Causal Games, a framework grounded in structural causal models that integrates multi-step strategic interaction with the Pearl Causal Hierarchy. Our model incorporates both the rational and the behavioral aspects of decision-making, explicitly distinguishes natural, interventional, and counterfactual actions, and strictly generalizes extensive-form games, repeated games, Markov games, and multi-agent influence diagrams. We develop counterfactual strategies, show they weakly dominate standard interventional strategies, and propose Causal Nash Equilibrium, a solution concept that endogenizes agents' choice of causal reasoning layer. Experiments in Kuhn poker and the iterated Prisoner's Dilemma also demonstrate that counterfactual agents can outperform classical game-theoretic strategies.

1. Introduction

Multi-step multi-agent reasoning lies at the heart of strategic interaction among decision-makers and plays a significant role across a wide range of social and technological domains, including medical treatment planning (Ning & Xie, 2024), financial markets, economic policymaking (Cui et al., 2022), robotic coordination and autonomous driving (Guestrin et al., 2001; Stone & Veloso, 2000; Zhang et al., 2024) and more recently agentic AI systems (Li et al., 2024; Guo et al., 2024). Motivated by these applications, sequential multi-agent decision-making has been studied extensively across economics, game theory and machine learning, giving rise to a broad family of models.

One of the earliest formal responses to this problem was the extensive-form game, introduced by Kuhn (1953). While the extensive-form representation was a foundational advance in making temporal order, information, and strategic commitment explicit, it suffers from the curse of dimensionality: the size of the game tree grows exponentially with the length of interaction, even for systems that admit compact descriptions under alternative representations. One such abstraction is provided by repeated games, introduced by Aumann (1959), which model interaction as the repeated play of a fixed-stage game over a finite or infinite horizon. However, both extensive-form games and repeated games lack any explicit abstraction of state or *mechanism*. As a result, they offer limited support for reasoning about structural dependencies, latent factors, or the effects of interventions beyond those explicitly enumerated in the game tree.

Stochastic games, also known as Markov games (Shapley, 1953), are able to mitigate the representational explosion of tree-based models by introducing a state-based abstraction (Littman, 1994; Busoniu et al., 2008). However, this abstraction comes at the cost of a strong Markovianity assumption. This requirement rules out the explicit representation of latent variables or hidden common causes that may influence both agent actions and state transitions across time. Moreover, the transition dynamics are specified only as conditional distributions, without encoding the structural relationships between state components, actions, and outcomes.

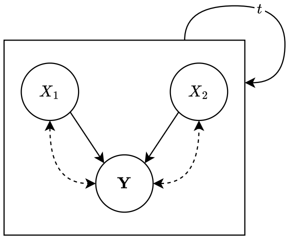
Graphical abstractions (Kearns et al., 2001; Vickrey & Koller, 2002) such as multi-agent influence diagrams (MAIDs) (Koller & Milch, 2003) attempt to address some of these limitations by exploiting conditional independence structure. Nevertheless, MAIDs remain fundamentally associational: edges represent probabilistic or informational dependence rather than causal mechanisms. While the framework specifies what information is available to each agent at each decision point, it provides no semantics for latent confounders or actions that span across the Pearl Causal Hierarchy (PCH): interventions and counterfactual reasoning. (Bareinboim et al., 2015; Maiti et al., 2025) has shown that such distinctions are not only an additional feature, but essential in incorporating aspects of both rational and irrational decision-making. Such shortcomings in all the prior models can be illustrated through the following example.

[1] Causal Artificial Intelligence Lab, Columbia University. Correspondence to: Aurghya Maiti <am5887@columbia.edu>.

Preprint. March 5, 2026.

		$X_2 = 0$		$X_2 = 1$	
		$U_2 = 0$	$U_2 = 1$	$U_2 = 0$	$U_2 = 1$
$X_1 = 0$	$U_1 = 0$	$-1.4, -1.4$	$-1, -1$	$-14, 0$	$-8, 0$
	$U_1 = 1$	$-1, -1$	$0, 0$	$0, 0$	$0, -3$
$X_1 = 1$	$U_1 = 0$	$0, -14$	$0, 0$	$0, 0$	$0, -8$
	$U_1 = 1$	$0, -8$	$-3, 0$	$-8, 0$	$0, 0$

(a)



(b)

Figure 1. (a) Y_1, Y_2 as functions of X_1, X_2, U_1, U_2; (b) Causal diagram of the iterated Prisoner's Dilemma.

P_1	P_2	M_1	M_2
TfT	TfT	-1.0, -1.0	-1.0, -1.0
TfT	D	-1.9, -1.9	-1.9, -1.9
TfT	Int	-3.56, -0.08	-1.56, -4.28
D	Int	-2.4, 0	0, -8.9

Figure 2. Payoffs under different strategies in the Iterated Causal Prisoner's Dilemma

Example 1.1 (Iterated Causal Prisoner's Dilemma)**.** Two friends repeatedly engage in criminal activities. After each offense, they are apprehended and questioned separately. Due to limited evidence, convictions depend on their strategic choices. In each episode $t \in \{1, \ldots, T\}$, each individual must decide whether to remain silent (cooperate, denoted by C) or betray the other (defect, denoted by D). Let $X_{i,t} \in \{0, 1\}$ denote the action of individual $i \in \{1, 2\}$ at episode t, where 0 corresponds to cooperation and 1 to defection.

Each episode is influenced by latent circumstances, represented by variables $U_{i,t} \in \{0, 1\}$. These variables capture unobserved factors such as the strength of evidence available to the prosecution, the competence of legal counsel, the disposition of the judge or jury, or the emergence of new witnesses. While these circumstances are not directly observable by the individuals, they affect both their behavior and outcomes. We assume that for each episode t and each individual i, circumstances are adversarial with probability $P(U_{i,t} = 0) = 0.6$.

Each of the prisoner possesses an inherent ability to assess their circumstances in each episode, denoted by $R_{i,t} \in \{0, 1\}$. When individual i accurately assesses their situation ($R_{i,t} = 1$), they cooperate if circumstances are favorable ($U_{i,t} = 1$) and defect otherwise. When their assessment is inaccurate ($R_{i,t} = 0$), this behavior is reversed. The resulting instinctive or behavioral action of individual i in episode t is given by the structural equation $X_{i,t} \leftarrow f_X(R_{i,t}, U_{i,t}) = R_{i,t} \oplus U_{i,t}$ where $\oplus$ denotes the exclusive-or operator. The latent variables $\{U_{i,t}, R_{i,t}\}_{t=1}^{T}$ and the function f_X are fixed and unknown to the individuals.

Given the realized actions and latent circumstances in episode t, the outcome $\mathbf{Y}_t = (Y_{1,t}, Y_{2,t})$ is generated according to the payoff structure shown in Fig. 1a. For example, when both individuals face favorable circumstances and cooperate, they incur minimal penalties, whereas asymmetric circumstances combined with defection can lead to severe punishment for the cooperating individual. The overall utility of each individual is defined as the mean of episode-level outcomes, $(1/T) \sum_{t=1}^{T} Y_{i,t}$.

We consider two environments that are indistinguishable from the perspective of intervention-level actions. In environment M_1, both individuals consistently assess their circumstances correctly, so that $R_{1,t} = R_{2,t} = 1$ for all t. In environment M_2, both individuals consistently misjudge their circumstances, so that $R_{1,t} = R_{2,t} = 0$ for all t. In both environments, the marginal distribution of $(U_{1,t}, U_{2,t})$ is identical across episodes.

If individuals ignore their intuition and optimize directly over actions at each episode, the induced stage game is identical in both environments and coincides with the classical Iterated Prisoner's Dilemma. Consequently, the Nash equilibrium strategies—such as always defect (D) and tit-for-tat (TfT)—are the same in both cases (Shoham & Leyton-Brown, 2008).

In contrast, when individuals follow their behavioral decision rules, long-run outcomes diverge sharply. In environment M_1, instinctive behavior yields an average per-episode payoff of $(0, 0)$, whereas in M_2 it results in approximately $(-2.4, -2.4)$ (see Appendix C). Moreover, when one agent follows a behavioral policy while the other plays a Nash equilibrium strategy such as tit-for-tat, the equilibrium player incurs a payoff of -3.56, whereas the behavioral agent attains -0.08. Figure 2 summarizes expected payoffs across tit-for-tat (TfT), always defect (D), and behavioral (Int) strategies under both environments. While tit-for-tat performs competitively within the interventional policy class, its performance degrades substantially against behavioral policies. Notably, although M_1 and M_2 are indistinguishable at the interventional level, they induce markedly different outcomes under behavioral reasoning, motivating the question of whether agents should commit to game-theoretic rationality or instead act according to behavioral policy.

In this work, we turn to a framework rooted in causal inference, which provides explicit semantics for mechanisms, interventions, and counterfactual reasoning (Pearl, 2009; Bareinboim, 2025). In single-agent settings, causal inference has been successfully applied to improve decision-making by formalizing agency and action, guiding where to intervene and what to observe, and enabling agents to act counterfactually (Bareinboim et al., 2015; 2024). Beyond structural questions, causal methods have also yielded substantial gains on statistical fronts, including reductions in sample complexity through transfer learning, principled use of offline data, and generalization across environments with shared causal structure. These advances suggest that causal reasoning provides exactly the representational and inferential tools missing from existing sequential game-theoretic models, motivating its extension to multi-agent, multi-step strategic settings.

Recent work by Maiti et al. (2025) makes a step towards integrating causal reasoning into multi-agent systems[1]. However, the proposed framework is limited to single-step interactions and does not account for the sequential structure that characterizes many real-world decision-making problems. In contrast, this work addresses the sequential setting, with the following contributions:

1. We formalize a class of games (Def. 2.8) that combine rational and behavioral aspects of decision-making and allows unobserved confounding and non-Markovianity. We then show this is a strictly richer representation than Extensive Form (Th. 2.9, 2.10) Games and Markov Games (Th. 2.11).
2. We introduce counterfactual strategies that are shown to be better than interventional strategies as studied in game theory (Th. 3.6).
3. We show the existence of Causal Nash Equilibrium and propose an algorithm for computing such equilibrium strategies.

In Sec . 4, we show experiments on Kuhn Poker and Iterated Prisoner's Dilemma to show that counterfactual strategies can be better in reality against interventional strategies.

1.1. Preliminaries

In this section, we introduce the notations and definitions used throughout the paper. We use capital letters to denote random variables (X) and small letters to denote their values (x). $\mathcal{D}_X$ denotes the domain of X. $|\mathbf{S}|$ denotes the cardinality of the set $\mathbf{S}$. $[n]$ denotes the set $\{1, 2, \ldots, n\}$, and $<$ or $>$ between tuples denote Pareto domination. The basic framework of our model resides on Structural Causal Models (Pearl, 2009). An SCM M is a tuple $\langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{U})\rangle$, where $\mathbf{V}$ and $\mathbf{U}$ are sets of endogenous and exogenous variables respectively. $\mathcal{F}$ is a set of functions f_V determining the value of $V \in \mathbf{V}$, that is, $V \leftarrow f_V(\mathbf{Pa}(V), \mathbf{U}_V)$, where $\mathbf{Pa}_V \subseteq \mathbf{V}$ and $\mathbf{U}_V \subseteq \mathbf{U}$. Naturally, M induces a distribution over the endogenous variables, $P(\mathbf{V})$, called *observational or L_1 distribution*. An intervention on a subset $\mathbf{X} \subseteq \mathbf{V}$, denoted by $do(\mathbf{x})$ is an operation where values of $\mathbf{X}$ are set to $\mathbf{x}$, replacing functions $\{f_X : X \in \mathbf{X}\}$, that would normally determine their values. For an SCM M, $M_\mathbf{x}$ denotes the model induced by $do(\mathbf{x})$ and $P(\mathbf{Y}_\mathbf{x})$ denotes the probability of $\mathbf{Y}$ in $M_\mathbf{x}$. Such distributions are called *interventional or L_2 distributions*. For more details, refer to Appendix A.1, Bareinboim et al. (2022) and Maiti et al. (2025).

2. Causal Games

In this section, we propose a new representation of sequential games, which we call Causal Games based on SCMs. The agents in Causal Games may be able to interact with the system through the different layers of the PCH and here, we define the system and the policy space corresponding to the three layers.

Definition 2.1 (Causal Multi-Agent System)**.** A Causal Multi-Agent System (CMAS) is a tuple $\langle M, N, \mathbf{X}, \mathbf{Y}\rangle$, where $M : \langle \mathbf{U}, \mathbf{V}, \mathcal{F}, \mathbb{P}\rangle$ is an SCM and

- N is the set of n agents,
- $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is the ordered set of action nodes with disjoint $\mathbf{X}_i, \mathbf{X}_j \subset \mathbf{V}$ for $i, j \in [n]$,
- $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ is the ordered set of reward signals, with $\mathbf{Y}_i \subseteq \mathbf{V}$ for all $i \in [n]$. □

A CMAS is essentially an SCM that contains nodes $\mathbf{X}$ that represent actions available to the n agents in the system. Each agent has control over a distinct subset of action nodes; so, no two agents can act on the same variable. Also, the system contains a set of reward variables, $\mathbf{Y}$, which represents the feedback or payoff that each agent receives based on their actions and the underlying causal mechanism. In order to illustrate some of the concepts introduced in this section, we will walk through the following variation of the popular sharing game studied in the literature (Shoham & Leyton-Brown, 2008).

Example 2.2 (Causal Sharing Game)**.** Two siblings receive a pair of Christmas gifts from their parents. The parents allow the siblings to decide how to share the gifts between

[1] The authors also discuss limitations of related approaches, including Hammond et al. (2023).

$f_Y(X_1, X_2, U_1, U_2)$		$X_2 = 0$		$X_2 = 1$	
		$U_2 = 0$	$U_2 = 1$	$U_2 = 0$	$U_2 = 1$
$X_1 =$ 2-0	$U_1 = 0$	1, 1	1, -1	-2, 1	-2, -1
	$U_1 = 1$	-1, 1	-1, -1	6, 1	6, -1
$X_1 =$ 1-1	$U_1 = 0$	1, 1	1, -1	-1, -1	-1, 3
	$U_1 = 1$	-1, 1	-1, -1	3, -1	3, 3
$X_1 =$ 0-2	$U_1 = 0$	1, 1	1, -1	1, -2	1, 6
	$U_1 = 1$	-1, 1	-1, -1	-1, -2	-1, 6

Figure 3. $\mathbf{Y} = (Y_1, Y_2)$ as a function of X_1, X_2, U_1, U_2 in the Causal Sharing Game

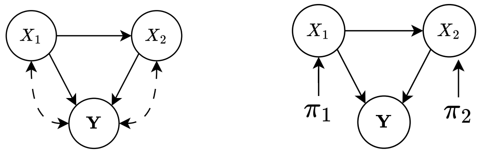
themselves. The interaction proceeds sequentially with perfect information. First, the older sibling (Agent 1) proposes a division of the gifts, after which the younger sibling (Agent 2) decides whether to accept or reject the proposal. If the proposal is accepted, the gifts are distributed as proposed; if rejected, neither of them get any gift. We denote the proposal of Agent 1 by the variable X_1, which takes values in a finite set of possible splits (e.g., 2-0, 1-1, or 0-2), representing the number of gifts allocated to Agent 1 and Agent 2, respectively. The response of Agent 2 is denoted by $X_2 \in \{0, 1\}$, where $X_2 = 1$ corresponds to acceptance and $X_2 = 0$ to rejection.

The enjoyment each sibling derives from the gifts depends not only on the number of gifts they receive but also on how much they personally like the specific gifts. Let the latent variables U_1 and U_2 denote how much the gifts follow their preferences. However, since the gifts are well-wrapped for surprise, they don't have any way of knowing what is inside. However, based on several factors, such as size/weight, which parents got the gift and so on, they may be able to get an intuition about whether the gift is good for them or not. In particular, Agent i has an internal perception variable R_i that reflects how accurately they judge whether the gift is upto their liking or not. If Agent 2 has an accurate reading of their preferences ($R_2 = 1$), they accept a proposed split when it aligns with their preference U_2, and reject it otherwise; if their reading is inaccurate ($R_2 = 0$), this behavior is reversed. Thus, Agent 2's instinctive response to a proposal is governed by a structural rule of the form $X_2 \leftarrow f_{X_2}(X_1, R_2, U_2)$, where the function f_{X_2} encodes how perceived and actual enjoyment jointly determine acceptance. The variables U_1, U_2, R_2, and the function f_{X_2} are fixed by nature and are unknown to the agents. For our example, let X_1 and X_2 are determined as following:

$$X_1 = U_1 \cdot (2, 0) + (1 - U_1) \cdot (0, 2) \tag{1}$$

$$X_2 = \mathbb{1}\{X_2 = (2, 0)\} + \mathbb{1}\{X_2 \neq (2, 0)\} \cdot U_2 \tag{2}$$

The realized outcome $\mathbf{Y} = (Y_1, Y_2)$, representing the enjoyment levels of the two siblings, is a function of the proposed split X_1, the response X_2, and the unobserved gifts (U_1, U_2), as illustrated in Fig. 3. For example, receiving fewer gifts that one strongly prefers may yield higher enjoyment than receiving more gifts that one does not like.



(a) *(b)*

Figure 4. (a) Causal Graph of the Sharing Game (b) The same causal graph under interventions

If both siblings ignore their intuitions and instead reason purely at the level of actions, the interaction reduces to the classical extensive-form sharing (or ultimatum) game. In this interventional view, the expected payoff for a proposal-response pair $(X_1 = x_1, X_2 = x_2)$ is given by

$$\sum_{u_1, u_2, \mathbf{y}} \mathbf{Y} \cdot P(u_1, u_2)\, P(\mathbf{Y} \mid x_1, x_2, u_1, u_2),$$

and equilibrium analysis proceeds without reference to the unobserved factors. This is represented by the interventional graph in Fig. 4a. Such an analysis may admit equilibria that appear fair or stable at the action level. However, when the siblings act according to the natural intuitions they get a payoff of $(3.5, 1.25)$ while the best they get under Nash Equilbrium is $(1, 1)$. □

To formalize these observations, we define different forms of actions that an agent may take in such a system and explore how they are related.

Definition 2.3 (L_1 action)**.** Given a CMAS $\langle N, M, \mathbf{X}, \mathbf{Y} \rangle$, an L_1 action of an agent i is the one in which the value of their action variables $\mathbf{X}_a$ are determined by the natural mechanism $f_{\mathbf{X}_i} \in \mathcal{F}$. □

We will also call such actions *behavioral actions* and denote them by a_0. Note that, while performing a_0, an agent does not know anything about the underlying SCM nor do they deliberately change any mechanism or variable in the system. The L_1 action space is thus $\mathcal{A}^1 = \{a_0\}$ and L_1 policy space is also a singleton set $\Pi^1 = \{a_0\}$.

Example 2.4. Consider the CMAS presented in Ex. 2.2. The L_1 action is when the values of X_1 and X_2 are determined by their natural function. The expected payoff when both the agents play behavioral action is given by

$$Y_1 = 0.5 \cdot 6 + 0.5 \cdot 1 = 3.5 \tag{3}$$

$$Y_2 = 0.25 \cdot (-1) + 0.25 \cdot 1 + 0.25 \cdot 6 = 1.25 \tag{4}$$

In a more traditional game theoretic sense, an agent can perform an intervention on the system. These interventions can be atomic interventions, where an agent sets the value of the action variable to a constant based on its context (Pearl, 2009), or soft interventions, where an agent samples their actions from a distribution (Correa & Bareinboim, 2020). Next, we define L_2 actions and the policy space.

Definition 2.5 (L_2-action)**.** Given a CMAS $\langle N, M, \mathbf{X}, \mathbf{Y}\rangle$, the L_2 action of an agent i is a sequence of mappings $\sigma_i = \{\sigma^i_1, \ldots, \sigma^i_H\}$ from states $\mathbf{S}_i = \{\mathbf{S}^i_1, \ldots \mathbf{S}^i_H\}$ to its action variables $\mathbf{X}_i = \{\mathbf{X}^i_1, \ldots, \mathbf{X}^i_H\}$ respectively, where

- Action $\mathbf{X}^i_j$ is a non-descendant of $\mathbf{X}^i_{j+1}, \ldots, \mathbf{X}^i_H$
- States $\mathbf{S}^i_j$ is a non-descendant of $\mathbf{X}^i_j, \ldots, \mathbf{X}^i_H$ □

Thus, if an agent i performs the action σ_i, then $\mathbf{X}^i_j$'s natural mechanism $f_{\mathbf{X}^i_j}$ is replaced by $\mathbf{X}^i_j \leftarrow \sigma_i(\mathbf{S}^i_j)$. The set of all such L_2 actions will be denoted by $\mathcal{A}^2$. L_2 policy is a distribution over the actions in $\mathcal{A}^2$.

Example 2.6. Consider the CMAS introduced in Ex. 2.2. L_2 action is when an agent performs an intervention, that is setting their action variable to a particular value. If Player 1 is playing 0-2 and Player 2 is playing Y only when Player 1 plays 0-2 and N otherwise, then the assignment of the variables are given by:

$$X_1 \leftarrow \text{0-2}, \quad X_2 \leftarrow \begin{cases} 1 & \text{if } X_1 = \text{0-2} \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

and U_1, U_2, R_1, R_2 are sampled from the distribution $P(\mathbf{U})$. Let's call this strategy σ_1, σ_2 and the expected payoff of this strategy is given by $(0, 2)$. It is also possible for one agent to perform an L_2 action and the other to perform an L_1 action.

In many cases, an agent can interact with the environment through PCH's Layer 3 (Bareinboim et al., 2015; 2022; Raghavan & Bareinboim, 2025). This allows agents to incorporate certain counterfactuals into their decision-making. For example, in Ex. 2.2, following natural instinct led to a suboptimal outcome for both agents. However, if both agents had done the exact opposite of their instinctive choices, they could have achieved a payoff of $(0, 0)$. This ability to override instinct and strategically adjust behavior falls within the realm of Layer 3 of PCH. Before formally defining L_3 actions, let $\mathbf{X}_i$ denote the action variable of the agent i, where its value is determined as a function f_i of its observable and unobservable parents $Pa^+(\mathbf{X}_a)$.

Definition 2.7 (L_3-action space $\mathcal{A}^3$)**.** Given a CMAS $\langle N, M, \mathbf{X}, \mathbf{Y}\rangle$, the L_2 action of an agent i is a sequence of mappings $\sigma_i = \{\sigma^i_1, \ldots, \sigma^i_H\}$ from states $\mathbf{S}_i = \{\mathbf{S}^i_1, \ldots \mathbf{S}^i_H\}$ and natural actions $\mathbf{X}'_i = \{\mathbf{X'}^i_1, \ldots, \mathbf{X'}^i_H\}$ to its realized action variables $\mathbf{X}_i = \{\mathbf{X}^i_1, \ldots, \mathbf{X}^i_H\}$ respectively, where

- Action $\mathbf{X}^i_j$ is a non-descendant of $\mathbf{X}^i_{j+1}, \ldots, \mathbf{X}^i_H$
- States $\mathbf{S}^i_j$ is a non-descendant of $\mathbf{X}^i_j, \ldots, \mathbf{X}^i_H$ □

When an agent takes an L_3 action, they first note their natural instinct $\mathbf{X}'_i$ and then make the decision $\mathbf{X}_i$ as follows:

$$\mathbf{X}'_i \leftarrow f_i(Pa^+(\mathbf{X}_i)), \quad \mathbf{X}^i_j \leftarrow \sigma_i(\mathbf{X'}^i_j, \mathbf{S}^i_j) \tag{6}$$

In case $h(x) = x$, it is the same as the natural or L_1 action, and if $h(x)$ is constant for all x, then it is an L_2 action. Bearing this in mind, we will often denote a_0 as $\mathbf{X} = \mathbf{X}'$, where $\mathbf{X}$ is the action variable and $\mathbf{X}'$ is the intuition.

Definition 2.8 (Causal Games)**.** A tuple $\Gamma = \langle \mathbb{M}, \mathcal{A}, \mathcal{R}\rangle$ is a Causal Normal Form Game (CNFG), where

- $\mathbb{M}$ is a CMAS $\langle N, M, \mathbf{X}, \mathbf{Y}\rangle$,
- $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_n)$ is the tuple of policies for the n agents, where $\mathcal{A}_i \in \{\mathcal{A}^1, \mathcal{A}^2, \mathcal{A}^1 \cup \mathcal{A}^2, \mathcal{A}^3\}$,
- $\mathcal{R} = (\mathcal{R}_1, \ldots, \mathcal{R}_n)$ is the tuple of reward functions, where $\mathcal{R}_i$ maps outcome $\mathbf{Y}_i$ to $\mathbb{R}$ for agent i. □

In the next section, we show that the causal games can be more expressive than several other widely used representations.

2.1. Extensive Normal Form Games

In this section, we relate Causal Games (CGs) to Extensive-Form Games (EFGs). We first show that CGs can offer a representational advantage over EFGs. We then argue that Causal Games capture semantic features that cannot be represented within the EFG formalism. For Nash Equilibrium to exist, we will assume the agents have perfect recall. A direct, naive conversion from an EFG to a CG may appear to incur an exponential increase in description length, since conditional probability distributions (CPDs) can be written as full tables over parent assignments. However, when CPDs exhibit context-specific independence and are represented compactly (e.g., as decision trees rather than tables) (Boutilier et al., 1996), the induced Causal Game description grows by at most a linear factor relative to the original EFG. Conversely, there exist families of extensive-form games whose explicit tree representations necessarily blow up exponentially, while the corresponding Causal Game admits a succinct description. Together, these observations yield the following two-part theorem.

Theorem 2.9 (Representation Gap Between EFGs and Causal Games)**.** *Let G be a finite EFG with perfect recall, and let $|G|$ denote the size of its standard description. Let Γ be a minimal causal game (CG) that is outcome-equivalent to G, and let $|\Gamma|$ denote the size of its description. Then the following hold:*

1. *For every finite EFG G, there exists an outcome-equivalent causal game Γ such that $|\Gamma| \leq O(|G|)$*
2. *There exists a family of finite EFGs $\{G_n\}_{n=1}^{\infty}$ and corresponding minimal outcome-equivalent causal games $\{\Gamma_n\}_{n=1}^{\infty}$ such that $|G_n| \geq 2^{\Omega(|\Gamma_n|)}$*

In words, this implies that converting an EFG into an outcome-equivalent CG incurs at most a linear increase in description length, while in some instances the CG representation is exponentially more succinct than the EFG. We next establish that CG constitute a strictly richer modeling framework than EFG.

Theorem 2.10. *Given any EFG, there exists two causal games Γ_1 and Γ_2, with L_1 payoffs μ_1 and μ_2 and Nash Equilibrium payoff μ_{NE}, such that $\mu_1 < \mu_{NE} < \mu_2$.* □

This result shows that two Causal Games can entail the same EFG under the interventional policy space, but can vary under the behavioral policy space.

2.2. Markov Games

In this section, we will look at another representation of sequential games, known as stochastic games or Markov Games (Shapley, 1953). Markov Games (MG) extend the notion by including the idea of state, which evolve based on the joint action of the agents. However, they still fail to capture the inherent hierarchical nature of actions, since actions in Markov Games can be interpreted as L_2 actions. A illustration of this with respect to MDP is provided in (Bareinboim et al., 2024), where two MDPs can coincide on the L_2 actions but differ in L_1 action space and vice versa. Such trends can also be observed here. Formally, the result can be stated in the following theorem.

Theorem 2.11. *Given any MG, there exists two causal games Γ_1 and Γ_2, with L_1 payoffs μ_1 and μ_2 and Nash Equilibrium payoff μ_{NE}, such that $\mu_1 < \mu_{NE} < \mu_2$.*

Another constraint of the Markov Games is the absence of unobserved confounding and the subsequent lack of unobserved confounders between states, policies, observations and reward functions.

3. Solving Causal Games

In this section, we discuss two solutions: Nash Equilibrium in the counterfactual space and the Causal Nash Equilibrium, which follows from Counterfactual Rationality proposed in (Maiti et al., 2025).

3.1. Nash Equlibrium in Causal Games

The first solution we propose is similar to finding the Nash Equilibrium in the whole space of counterfactual actions.

```
Algorithm 1 Find-CNE (Sequential)
Input: Sequential causal game $\Gamma$ and admissible PCH policy classes $\{\Pi_i^1, \Pi_i^2, \Pi_i^3\}_{i \in N}$
Output: CNE policy profile $\pi^*$
Construct the layer selection game $L_\Gamma$ by evaluating, for each $(\Pi_1, \ldots, \Pi_n)$, the Nash equilibrium payoff of $\Gamma(\Pi_1, \ldots, \Pi_n)$
Compute a Nash equilibrium $s^*$ of $L_\Gamma$
Define $\Pi_i^* = \bigcup_{\Pi_i \in \mathrm{supp}(s_i^*)} \Pi_i$ for each agent $i$
Compute a Nash equilibrium $\pi^*$ of the projected sequential game $\Gamma(\Pi_1^*, \ldots, \Pi_n^*)$
Return: $\pi^*$
```

Recall that a strategy of an agent i is a tuple of all policies $(\sigma_1^i, \ldots, \sigma_H^i)$ for determining all the action variables. In this section, we assume that the agents have perfect recall, that is, they are able to remember the actions taken previously. Similar to the standard literature, for a joint policy σ_{-i}, the best response σ_i^* is such that, $E[Y^i_{\sigma_i^*, \sigma_{-i}}] \geq E[Y^i_{\sigma_i', \sigma_{-i}}]$ for all σ_i'. Counterfactual best response is when the elements of σ_i belong to the set of counterfactual policy space. Since, the counterfactual action space is a superset of the interventional action space, we can write the following theorem.

Theorem 3.1. *The best response from the counterfactual policy space is as good as the response from L_2 space for any strategy profile of the other agents.*

Following the definition of best response, we can define Nash Equilibrium in the larger action space.

Definition 3.2 (Nash Equilibrium in Counterfactual Space)**.** Nash Equilibrium in Counterfactual Space is when all the players are playing the counterfactual best response.

As a corolllary to the existence of Nash Equilibrium, it is easy to show that the Nash Equilibrium in the counterfactual space also exists.

3.2. Causal Nash Equilibrium

We introduce a *layer selection metagame* for sequential Causal Games, capturing the fact that agents may act through different levels of the PCH. These levels induce distinct policy classes, ranging from natural (behavioral) policies to interventional and counterfactual policies. The key question is whether agents should always employ the most expressive policy class, or whether restricting attention to a subset of causal policies can be strategically optimal.

Formally, consider a finite-horizon sequential causal game induced by a structural causal model M, with agent set $N = \{1, \ldots, n\}$. Let $X_{i,t}$ and $Y_{i,t}$ denote the action and reward variables of agent i at time $t \in [T]$. For each agent i, let Π_i^1, Π_i^2, and Π_i^3 denote the sets of admissible policies corresponding to the first, second, and third layers of the

PCH, respectively. We also allow unions of layers, such as $\Pi_i^1 \cup \Pi_i^2$, to model agents that can flexibly switch between reasoning modes within a restricted subset of the hierarchy. Given a sequential causal game Γ, we define a *PCH projection* of Γ by restricting each agent's policy space to a chosen subset of layers.

Definition 3.3 (PCH Projection of CG)**.** Given a sequential causal game Γ and a profile of policy spaces $(\Pi_1, \dots, \Pi_n)$ with $\Pi_i \subseteq \{\Pi_i^1, \Pi_i^2, \Pi_i^1 \cup \Pi_i^2, \Pi_i^3\}$, the PCH projection of Γ, denoted by $\Gamma(\Pi_1, \dots, \Pi_n)$, is the sequential game obtained by restricting agent i to policies in Π_i, while keeping the underlying causal model and reward structure unchanged.

This projection captures how the strategic interaction evolves when agents are constrained to reason within specific layers of the causal hierarchy. Importantly, different projections of the same underlying game may induce different equilibria and different cumulative payoffs. The key problem is therefore to determine which projection is stable, in the sense that no agent has an incentive to unilaterally switch to a different layer of reasoning. To address this, we define a meta-game in which agents choose their reasoning layers before playing the underlying sequential game.

Definition 3.4 (Layer Selection Game)**.** Given a sequential causal game Γ, its layer selection game L_Γ is a normal-form game with the same set of agents N. The action set of agent i in L_Γ consists of admissible policy spaces $\Pi_i \in \{\Pi_i^1, \Pi_i^2, \Pi_i^1 \cup \Pi_i^2, \Pi_i^3\}$. For any joint selection $(\Pi_1, \dots, \Pi_n)$, the payoff to agent i in L_Γ is defined as the Nash equilibrium payoff of the projected sequential game $\Gamma(\Pi_1, \dots, \Pi_n)$, where equilibrium is computed with respect to policies over time.

Intuitively, each cell of the payoff matrix of L_Γ corresponds to a sequential game in which agents are restricted to specific layers of causal reasoning, and the payoff reflects the long-run utility achieved when agents play an equilibrium of that restricted game. We assume that the structure of the causal model and the policy classes are common knowledge, so that agents can reason about the consequences of selecting different layers.

Let s^* denote a Nash equilibrium of the layer selection game L_Γ. The support of s_i^* identifies the policy spaces that agent i assigns positive probability to. An agent may therefore rationally exclude policy spaces that are not in the support of s_i^*, effectively "forgetting" certain layers of reasoning if they are not optimal at the meta level.

Definition 3.5 (Causal Nash Equilibrium)**.** Let s^* be a Nash equilibrium of the layer selection game L_Γ, and define the induced policy space for agent i as $\Pi_i^* = \bigcup_{\Pi_i \in \mathrm{supp}(s_i^*)} \Pi_i$ A joint policy profile π^* is called a *Causal Nash Equilibrium (CNE)* of the sequential causal game Γ if π^* is a Nash equilibrium of the projected game $\Gamma(\Pi_1^*, \dots, \Pi_n^*)$.

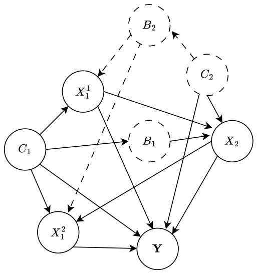


Figure 5. Causal Diagram for Kuhn Poker

In a CNE, no agent can improve its expected cumulative reward by unilaterally deviating either within the selected class of causal policies or by switching to a different reasoning layer that was excluded by the equilibrium of the layer selection game. Thus, CNE captures strategic stability both at the level of action generation over time and at the level of reasoning-layer choice.

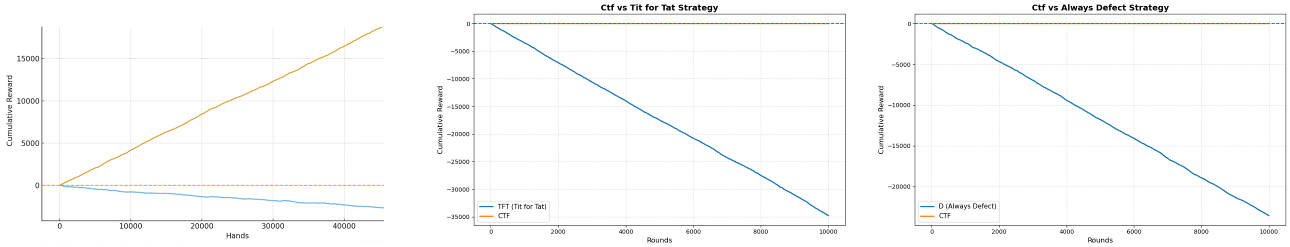
Existence of CNE. For any finite-horizon sequential causal game Γ, a Causal Nash Equilibrium always exists. This follows from the existence of a Nash equilibrium in the finite layer selection game L_Γ, together with the existence of equilibrium policies in each projected sequential game $\Gamma(\Pi_1, \dots, \Pi_n)$ under standard assumptions such as perfect recall and compact policy spaces. The definition of CNE naturally yields a constructive procedure for identifying equilibrium strategies. Finally, we establish a dominance property that clarifies the strategic meaning of the layer selection equilibrium.

Theorem 3.6 (Dominance of causal strategies)**.** *Let Γ be a finite-horizon sequential causal game with Causal Nash Equilibrium (CNE) payoff μ^*, and let L_Γ denote its layer selection game with Nash equilibrium strategy s^*. Suppose that s^* is a pure-strategy Nash equilibrium of L_Γ, and let $\Pi_i^* = \mathrm{supp}(s_i^*)$ denote the set of policy classes selected by agent i. Then, for every agent i and for every admissible policy class $\Pi_i \notin \Pi_i^*$, the expected cumulative reward obtained by agent i under the CNE satisfies*

$$\mu_i^* \;\geq\; \mathrm{NE}\big(\Gamma(\Pi_i, \Pi_{-i}^*)\big)\,,$$

where $\mathrm{NE}(\Gamma(\Pi_i, \Pi_{-i}^*))$ denotes the Nash equilibrium payoff of the projected sequential causal game in which agent i is restricted to policy class Π_i while all other agents are restricted to Π_{-i}^*.

This result establishes that when the layer selection game admits a pure-strategy equilibrium, no agent benefits from unilaterally switching to a different layer of the Pearl Causal Hierarchy in the underlying sequential interaction. Even when using the whole counterfactual action space is not beneficial, a two-level approach may still result in a better equilibrium than the Nash Equilibrium strategy.



(a) Kuhn poker: CTF vs. GTO

(b) IPD: CTF vs. TfT

(c) IPD: CTF vs. Defect

Figure 6. Cumulative rewards in sequential games. Counterfactual agents (orange) outperform classical strategies across Kuhn poker and the iterated Prisoner's Dilemma.

4. Experiments

4.1. Kuhn Poker

We evaluate the strategic value of counterfactual reasoning in Kuhn poker, a canonical imperfect-information sequential game with a well-characterized Nash equilibrium. The game is played by two players using a three-card deck (Jack, Queen, King): each player antes one chip, receives one private card, and engages in a single round of betting consisting of check/bet, call/fold decisions, with the higher card winning the pot at showdown. Classical equilibrium strategies condition only on the public betting history and assume that all strategically relevant information is captured by the formal game tree. In contrast, real-world play often involves information leakage through behavioral cues (e.g., timing or body language). To model this, we augment Kuhn poker with latent causal variables that influence players' natural behavior and are partially revealed through intuition-based signals. Counterfactual agents are allowed to condition their realized actions on these intuitions, while game-theoretic agents are restricted to standard interventional L_2 strategies.

The causal diagram for this game from the perspective of Player 1 is shown in Fig. 5. C_1 and C_2 are the cards received by the Player 1 and 2 respectively. X_1^1 and X_1^2 are the decisions of player 1 and X_2 is the decision of player 2. C_i may effect the body language of Player i, denoted by B_i which may effect the other players intuition.

When roles are alternated over repeated play, counterfactual agents achieve strictly positive expected value, while game-theoretic agents incur losses; Fig. 6a shows that cumulative rewards grow monotonically in favor of counterfactual agents. These results demonstrate that counterfactual strategies can systematically exploit information embedded in natural behavior that is invisible at the interventional level, even in games with well-understood equilibria.

4.2. Iterated Prisoner's Dilemma

We analyze the performance of tit-for-tat in the iterated Prisoner's Dilemma under the Causal Prisoner's Dilemma (CPD) framework, where agents may follow either standard interventional strategies or intuition-based (natural-action) strategies. Our results highlight a sharp asymmetry in the robustness of tit-for-tat across different classes of opponent behavior. Consistent with classical results, tit-for-tat performs well against conventional level-L_2 strategies such as always-defect and tit-for-tat itself. When played against these strategies, tit-for-tat yields relatively stable outcomes, including mutual cooperation when facing itself and bounded losses when facing defect. This aligns with its well-known reputation as a resilient strategy in standard iterated Prisoner's Dilemma settings.

However, this robustness breaks down when tit-for-tat is played against counterfactual agents that systematically act contrary to their natural or intuitive responses in the CPD environment. In this case, tit-for-tat performs substantially worse than even unconditional defection. The comparison of the strategies are shown in the Fig. 6b and Fig. 6c.

5. Conclusions

We introduced Causal Games (Def. 2.8), a causal extension of sequential game theory grounded in SCMs that explicitly distinguishes behavioral, interventional, and counterfactual actions. In Sec. 2, we showed that this framework strictly generalizes classical representations, admits a well-defined solution concept via Causal Nash Equilibrium (Sec. 3), and that counterfactual strategies weakly dominate standard game-theoretic strategies. Experiments in Kuhn poker and the iterated Prisoner's Dilemma illustrate that accounting for causal structure and counterfactual reasoning can yield substantial strategic advantages in sequential interaction.

6. Impact Statements

This paper presents work whose goal is to advance the field of machine learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Aumann, R. J. Acceptable points in general cooperative n-person games. *Contributions to the Theory of Games (AM-40)*, 4:287, 1959.

Bareinboim, E. Causal artificial intelligence: A roadmap for building causally intelligent systems. 2025.

Bareinboim, E., Forney, A., and Pearl, J. Bandits with unobserved confounders: A causal approach. *Advances in Neural Information Processing Systems*, 28, 2015.

Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. *On Pearl's Hierarchy and the Foundations of Causal Inference*, pp. 507–556. Association for Computing Machinery, New York, NY, USA, 1 edition, 2022. ISBN 9781450395861. URL https://doi.org/10.1145/3501714.3501743.

Bareinboim, E., Zhang, J., and Lee, S. An introduction to causal reinforcement learning. Technical Report R-65, Causal Artificial Intelligence Lab, Columbia University, Dec 2024. https://causalai.net/r65.pdf.

Boutilier, C., Friedman, N., Goldszmidt, M., and Koller, D. Context-specific independence in bayesian networks. In *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, UAI'96, pp. 115–123, San Francisco, CA, USA, 1996. Morgan Kaufmann Publishers Inc. ISBN 155860412X.

Busoniu, L., Babuska, R., and De Schutter, B. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2):156–172, 2008.

Correa, J. and Bareinboim, E. A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 10093–10100, 2020.

Cui, K., Tahir, A., Ekinci, G., Elshamanhory, A., Eich, Y., Li, M., and Koeppl, H. A survey on large-population systems and scalable multi-agent reinforcement learning. *arXiv preprint arXiv:2209.03859*, 2022.

Guestrin, C., Koller, D., and Parr, R. Multiagent planning with factored mdps. *Advances in neural information processing systems*, 14, 2001.

Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., and Zhang, X. Large language model based multi-agents: A survey of progress and challenges. In Larson, K. (ed.), *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pp. 8048–8057. International Joint Conferences on Artificial Intelligence Organization, 8 2024. doi: 10.24963/ijcai.2024/890. URL https://doi.org/10.24963/ijcai.2024/890. Survey Track.

Hammond, L., Fox, J., Everitt, T., Carey, R., Abate, A., and Wooldridge, M. Reasoning about causality in games. *Artificial Intelligence*, 320:103919, 2023.

Kearns, M., Littman, M. L., and Singh, S. Graphical models for game theory. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, pp. 253–260, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. ISBN 1558608001.

Koller, D. and Milch, B. Multi-agent influence diagrams for representing and solving games. *Games and economic behavior*, 45(1):181–221, 2003.

Kuhn, H. W. Extensive games and the problem of information. *Contributions to the Theory of Games*, 2(28): 193–216, 1953.

Li, X., Wang, S., Zeng, S., Wu, Y., and Yang, Y. A survey on llm-based multi-agent systems: workflow, infrastructure, and challenges. *Vicinagearth*, 1(1):9, 2024.

Littman, M. L. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, pp. 157–163. Elsevier, 1994.

Maiti, A., Jain, P., and Bareinboim, E. Counterfactual rationality: A causal approach to game theory. Technical Report R-125, Causal Artificial Intelligence Lab, Columbia University, USA, January 2025.

Ning, Z. and Xie, L. A survey on multi-agent reinforcement learning and its application. *Journal of Automation and Intelligence*, 3(2):73–91, 2024.

Pearl, J. *Causality*. Cambridge university press, 2009.

Raghavan, A. and Bareinboim, E. Counterfactual realizability and decision-making. In *The 13th International Conference on Learning Representations*, 2025. forthcoming.

Shapley, L. S. Stochastic games. *Proceedings of the national academy of sciences*, 39(10):1095–1100, 1953.

Shoham, Y. and Leyton-Brown, K. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

Stone, P. and Veloso, M. Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, 8 (3):345–383, 2000.

Vickrey, D. and Koller, D. Multi-agent algorithms for solving graphical games. *AAAI/IAAI*, 2:345–351, 2002.

Zhang, R., Hou, J., Walter, F., Gu, S., Guan, J., Röhrbein, F., Du, Y., Cai, P., Chen, G., and Knoll, A. Multi-agent reinforcement learning for autonomous driving: A survey. *arXiv preprint arXiv:2408.09675*, 2024.

A. Preliminaries and Background

A.1. Structural Causal Models and the Pearl Causal Hierarchy

Structural Causal Models (SCMs) provide a unifying framework for representing data-generating processes under explicit causal assumptions (Pearl, 2009; Bareinboim et al., 2024). A defining feature of SCMs is that they support multiple modes of interaction with a system, ranging from passive observation to active intervention and counterfactual reasoning. These modes form the *Pearl Causal Hierarchy* (PCH), which organizes causal queries by increasing expressive power. Our presentation follows Bareinboim et al. (2022).

Definition A.1 (Structural Causal Model)**.** A structural causal model is a tuple $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}) \rangle$, where $\mathbf{U}$ is a set of exogenous (background) variables, $\mathbf{V} = \{V_1, \dots, V_n\}$ is a set of endogenous variables, $\mathcal{F} = \{f_1, \dots, f_n\}$ is a collection of structural assignments $V_i \leftarrow f_i(Pa_i, U_i)$ with $Pa_i \subseteq \mathbf{V} \setminus \{V_i\}$ and $U_i \subseteq \mathbf{U}$, and $P(\mathbf{U})$ is a joint distribution over the exogenous variables.

The qualitative structure of an SCM is represented by a causal diagram, which encodes functional dependencies and unobserved confounding.

Definition A.2 (Causal Diagram)**.** Given an SCM $\mathcal{M}$, its causal diagram is a graph whose nodes correspond to endogenous variables, with a directed edge $V_i \to V_j$ if V_i appears in f_j, and a bidirected edge $V_i \leftrightarrow V_j$ if V_i and V_j share a common exogenous cause or have statistically dependent exogenous variables.

SCMs induce three families of probability distributions corresponding to the levels of the Pearl Causal Hierarchy.

Definition A.3 (L_1 valuation (Observational))**.** The observational distribution induced by $\mathcal{M}$ over $\mathbf{Y} \subseteq \mathbf{V}$ is $P^{\mathcal{M}}(\mathbf{y}) = \sum_{\mathbf{u}:\mathbf{Y}(\mathbf{u})=\mathbf{y}} P(\mathbf{u})$, where $\mathbf{Y}(\mathbf{u})$ denotes the solution of the structural equations under exogenous realization $\mathbf{u}$.

Interventional distributions are defined by modifying the structural equations.

Definition A.4 (L_2 valuation (Interventional))**.** For an intervention $\mathbf{X} \leftarrow \mathbf{x}$, the interventional distribution over $\mathbf{Y}$ is $P^{\mathcal{M}}(\mathbf{y_x}) = \sum_{\mathbf{u}:\mathbf{Y_x}(\mathbf{u})=\mathbf{y}} P(\mathbf{u})$, where $\mathbf{Y_x}(\mathbf{u})$ denotes the value of $\mathbf{Y}$ in the intervened model.

Finally, SCMs support counterfactual reasoning, which compares hypothetical outcomes for the same individual under incompatible interventions.

Definition A.5 (L_3 valuation (Counterfactual))**.** For $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, $P^{\mathcal{M}}(\mathbf{y_x}, \mathbf{z_w}) = \sum_{\mathbf{u}:\mathbf{Y_x}(\mathbf{u})=\mathbf{y}, \mathbf{Z_w}(\mathbf{u})=\mathbf{z}} P(\mathbf{u})$.

Together, the L_1, L_2, and L_3 valuations constitute the Pearl Causal Hierarchy, enabling increasingly expressive forms of causal reasoning.

A.2. Counterfactual Randomization

Although SCMs formally support counterfactual reasoning, interacting with a system at the counterfactual (L_3) level presents practical challenges. In particular, an agent's internal deliberation may involve considering and discarding multiple alternatives before committing to a final action, which raises ambiguity about which choice should be treated as the agent's natural intention. Counterfactual randomization resolves this ambiguity by identifying the agent's final intended action immediately prior to execution as the relevant intuition and randomizing the executed action conditionally on this intention (Bareinboim et al., 2015; 2024).

Concretely, the agent is interrupted just before acting, the intended action X is recorded, and an action X' is selected at random for execution. This procedure enables estimation of counterfactual quantities of the form $\mathbb{E}[Y_{X \leftarrow x} \mid X = x']$, corresponding to outcomes under interventions conditioned on the agent's natural inclination.

Figure 7 illustrates the decision flow underlying counterfactual randomization. For further details in the single-agent setting, see Bareinboim et al. (2024, Sec. 7).

A. Proofs

A.1. Proof of Theorem 2.10 (Representation Gap)

Proof. We fix a conventional encoding of finite EFGs and SCMs in which description length is Θ(#nodes + #edges + #tables/trees for local kernels + payoff data). All such encodings are equivalent up to constant factors for asymptotics.

Algorithm 2 `Ctf-RCT`: Counterfactual Randomized Controlled Trials

```
Input: action domain $\mathcal{D}(X)$, number of trials $N$
for $t = 1, 2, \dots$ do
    Observe intended action $X^{(t)}$
    if $t \leq N$ then
        Sample $X'^{(t)} \sim \mathrm{Unif}(\mathcal{D}(X))$
    else
        $X'^{(t)} \leftarrow \arg\max_x \widehat{\mathbb{E}}^{(N)}[Y_{X \leftarrow x} \mid X = X^{(t)}]$
    end if
    Execute $\mathrm{do}(X'^{(t)})$ and observe $Y^{(t)}$
end for
```

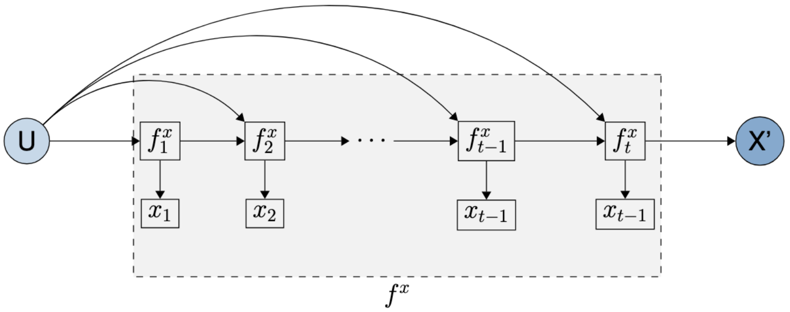


Figure 7. Illustration of decision flow f_X

Part (1). Let G be a finite EFG with perfect recall. We construct an SCM M_G that simulates play in G.

Let H be the set of nonterminal histories and Z the set of terminal histories. For each nonterminal history $h \in H$, let $P(h) \in N \cup \{c\}$ denote the acting player (or chance), and let $A(h)$ be the finite action set available at h. If $P(h) = c$, the EFG specifies a distribution $p_h(\cdot)$ on $A(h)$.

Introduce endogenous variables: (1) For each decision history h with $P(h) = i \in N$, an action variable $X_h \in A(h)$. (2) A terminal variable $S \in Z$ encoding the realized terminal history. Introduce exogenous variables: For each chance history h with $P(h) = c$, an exogenous draw $U_h \in A(h)$ with $\mathbb{P}(U_h = a) = p_h(a)$.

Define S as the (deterministic) result of traversing the tree from the root: at a player node h, choose the successor determined by X_h; at a chance node h, choose the successor determined by U_h; stop at the first terminal z and set $S = z$. This can be implemented via intermediate depth-indexed history variables H_t; acyclicity holds because the tree is finite and traversal proceeds forward in depth.

For each player i, define $Y_i := u_i(S)$ where $u_i : Z \to \mathbb{R}$ is the EFG payoff.

Because G has perfect recall, mixed (behavioral) strategies suffice. For each player i, map any behavioral strategy $\sigma_i(\cdot \mid I)$ (for information set I) to an L_2 policy that intervenes on each X_h with $P(h) = i$ by sampling X_h according to $\sigma_i(\cdot \mid I(h))$. Under this mapping, the induced distribution over S matches the distribution over terminal histories in G under σ, since chance moves are matched by U_h and player choices are matched by the interventions on X_h. Therefore expected payoffs match.

Total Size: The number of variables and local mechanisms is linear in the number of nodes/edges of the tree (up to constant-factor overhead for implementing S with intermediate H_t variables). If conditional distributions are represented by decision trees (context-specific independence), the representation is $O(|G|)$. Hence $|\Gamma| \leq O(|G|)$.

Part (2). We exhibit a family with exponential blow-up for EFG descriptions.

Fix a branching factor $b \geq 2$ and consider a finite-horizon n-step interaction with a compact state description (or even a stage game repeated n times where payoffs depend only on a small state). As an extensive-form game, the explicit tree must represent all action histories, yielding $\Omega(b^n)$ histories/nodes and hence $|G_n| = \Omega(b^n)$ under standard encodings.

As a causal game, represent the same interaction with variables $\{X_{i,t}\}_{i\in N,\, t\in[n]}$ and a small state variable S_t with a time-homogeneous transition equation $S_{t+1} := f(S_t, X_{1,t}, \dots, X_{|N|,t}, U_t)$ and reward equations $Y_{i,t} := g_i(S_t, X_{1,t}, \dots, U_t)$. This SCM has $O(n)$ variables and $O(n)$ equations with constant-size descriptions, so $|\Gamma_n| = O(n)$. Therefore $|G_n| = \Omega(b^n) = 2^{\Omega(n)} = 2^{\Omega(|\Gamma_n|)}$.

One concrete instance is a two-agent finite-horizon Markov interaction with binary variables. Let $S_t \in \{0,1\}$ be a state and $X_{1,t}, X_{2,t} \in \{0,1\}$ be the agents' actions for $t = 0, \dots, n-1$. At each step, rewards satisfy $Y_{i,t} \in \{0,1\}$ with $P(Y_{i,t} \mid S_t, X_{1,t}, X_{2,t})$, and the state evolves according to a time-homogeneous transition kernel $P(S_{t+1} \mid S_t, X_{1,t}, X_{2,t})$. Since each local CPD has a constant number of parameters (binary parents), the resulting causal-game/SCM description has size $O(n)$ in the horizon. In contrast, the explicit extensive-form tree must enumerate all action histories, yielding $\Omega(2^n)$ histories/nodes and thus exponential description length. □

A.2. Proof of Theorem 2.10 (EFG richness)

Proof. Fix an extensive-form game (EFG) G with finite horizon, finite player set $N = \{1, \dots, n\}$, and bounded terminal utilities $u_i(z) \in [-U, U]$ for every terminal history z and every $i \in N$. Let μ_{NE} denote the payoff vector of *some* Nash equilibrium of G (in behavioral strategies; existence is standard for finite EFGs).

Step 1: A baseline causal game whose L_2-projection is exactly the EFG. We construct a causal game Γ^0 induced by an SCM that reproduces the EFG under interventions. Index decision points by information sets: for each player i, let $\mathcal{I}_i$ be the set of information sets of i, and let $A(I)$ be the finite action set available at $I \in \mathcal{I}_i$. (Chance nodes may be treated as belonging to a dummy player "0" with fixed stochastic policy; this does not affect the argument.)

For each $I \in \mathcal{I}_i$, introduce an action variable $X_I \in A(I)$. Introduce additional endogenous variables encoding the realized history (or equivalently, the realized node) as the play unfolds; denote this collection by H. Let Z denote the terminal history variable, and define payoffs $Y_i := u_i(Z)$.

The SCM contains:

- structural equations mapping the collection of chosen actions $(X_I)_I$ and chance outcomes into a unique terminal history Z (this is well-defined because an EFG induces a deterministic successor relation given actions and chance outcomes);
- payoff equations $Y_i \leftarrow u_i(Z)$.

Under L_2 reasoning, agents intervene on their action variables at each information set, i.e., they replace the natural mechanism of X_I by a chosen (behavioral) decision rule. Thus, the induced L_2 game of Γ^0 is behaviorally equivalent to the original EFG G. In particular, the set of Nash equilibria and their payoff vectors coincide; hence there exists an L_2 Nash equilibrium payoff vector equal to μ_{NE}.

Step 2: Add L_1 mechanisms and "mean-zero under intervention" bonus terms. We now modify Γ^0 to obtain a one-parameter family of causal games $\Gamma(\alpha)$, where $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$. For each player i and each information set $I \in \mathcal{I}_i$, introduce an exogenous variable U_I taking values in $A(I)$. Fix, for each $I \in \mathcal{I}_i$, a distribution q_I over $A(I)$ with *full support* (e.g., uniform), and set $U_I \sim q_I$ mutually independent across I. Define the *natural (L1) action mechanism* as

$$X_I \leftarrow U_I \qquad \text{for every information set } I. \tag{7}$$

(Under interventions, X_I can be set arbitrarily as usual.)

Next, modify payoffs by adding a bonus term that vanishes in expectation under interventions. For each player i, define

$$Y_i \;:=\; u_i(Z) \;+\; \alpha_i \sum_{I\in\mathcal{I}_i} \Big(\mathbf{1}\{U_I = X_I\} - q_I(X_I)\Big), \tag{8}$$

where $q_I(X_I)$ denotes the probability mass that q_I assigns to the realized action X_I.

Lemma 1 (Interventional invariance). For any (behavioral) intervention profile on the action variables $(X_I)_I$,

$$\mathbb{E}[\mathbf{1}\{U_I = X_I\} - q_I(X_I) \mid \mathrm{do}(X_I = a)] \;=\; 0 \qquad \text{for every } I.$$

Proof. Condition on $\mathrm{do}(X_I = a)$. Then

$$\mathbb{E}[\mathbf{1}\{U_I = X_I\} \mid \mathrm{do}(X_I = a)] = \Pr(U_I = a) = q_I(a),$$

and $q_I(X_I) = q_I(a)$ deterministically under the intervention. Subtracting yields 0. ■

By linearity of expectation, Lemma 1 implies that for every intervention profile,

$$\mathbb{E}[Y_i \mid \mathrm{do}(\cdot)] \;=\; \mathbb{E}[u_i(Z) \mid \mathrm{do}(\cdot)]. \tag{9}$$

Therefore, the L_2-induced game of $\Gamma(\alpha)$ is *identical* to the original EFG G (same expected utilities for every strategy profile), and hence it admits a Nash equilibrium with payoff vector μ_{NE}. In particular, the Nash equilibrium payoff of $\Gamma(\alpha)$ equals μ_{NE}, independent of α.

Lemma 2 (Strict L_1 shift). Under L_1 play (i.e., with the natural mechanisms (7) and no interventions), for every information set I we have $X_I = U_I$ almost surely, and thus

$$\mathbb{E}\big[\mathbf{1}\{U_I = X_I\} - q_I(X_I)\big] = 1 - \sum_{a \in A(I)} q_I(a)^2 \;>\; 0,$$

where strict positivity holds because q_I has full support and $|A(I)| \geq 2$ for at least one information set of each player[2].

Proof. Under L_1, $X_I = U_I$ almost surely, so $\mathbf{1}\{U_I = X_I\} = 1$ a.s. Also, $X_I \sim q_I$, hence $\mathbb{E}[q_I(X_I)] = \sum_a q_I(a)^2$. The claimed expression follows. ■

Let

$$\Delta_i \;:=\; \sum_{I \in \mathcal{I}_i} \Big(1 - \sum_{a \in A(I)} q_I(a)^2\Big) \;>\; 0.$$

Write the L_1 payoff vector of $\Gamma(\alpha)$ as $\mu^{L_1}(\alpha)$. By (8) and Lemma 2,

$$\mu_i^{L_1}(\alpha) \;=\; \mathbb{E}[u_i(Z) \mid L_1] \;+\; \alpha_i \Delta_i, \qquad i \in N. \tag{10}$$

Step 3: Choose α to sandwich μ_{NE}. Since $u_i(z) \in [-U, U]$ and the horizon is finite, $\mathbb{E}[u_i(Z) \mid L_1] \in [-U, U]$ for all i. Fix any scalar $K > 0$ such that

$$K\Delta_i \;>\; U + |\mu_{\mathrm{NE},i}| \qquad \text{for all } i \in N.$$

Define α^+ by $\alpha_i^+ = +K$ for all i, and α^- by $\alpha_i^- = -K$ for all i. Let $\Gamma_2 := \Gamma(\alpha^+)$ and $\Gamma_1 := \Gamma(\alpha^-)$. Let μ_2 and μ_1 denote their respective L_1 payoff vectors, and recall that both have Nash equilibrium payoff μ_{NE} by (9).

From (10) and the choice of K,

$$\mu_{1,i} = \mathbb{E}[u_i(Z) \mid L_1] - K\Delta_i < -|\mu_{\mathrm{NE},i}| \leq \mu_{\mathrm{NE},i},$$

and similarly

$$\mu_{2,i} = \mathbb{E}[u_i(Z) \mid L_1] + K\Delta_i > |\mu_{\mathrm{NE},i}| \geq \mu_{\mathrm{NE},i}.$$

Hence, componentwise,

$$\mu_1 \;<\; \mu_{\mathrm{NE}} \;<\; \mu_2.$$

This completes the proof. □

[2] If a player has only singleton action sets at all information sets, then that player is strategically trivial and can be ignored; the strict inequalities can be enforced for all nontrivial players.

A.3. Proof of Theorem 2.11 (MG richness)

Proof. Fix an arbitrary (finite-state, finite-action) Markov game

$$G \;=\; \big\langle N,\, \mathcal{S},\, (\mathcal{A}_i)_{i\in N},\, P(\cdot \mid s,a),\, (r_i(s,a))_{i\in N},\, \rho_0,\, \gamma \big\rangle,$$

with discount factor $\gamma \in (0,1)$ and initial-state distribution ρ_0.[3] Let $\pi^\star = (\pi_1^\star, \ldots, \pi_n^\star)$ be a (stationary) Nash equilibrium of G with value (equilibrium payoff) vector μ_{NE}.[4]

Step 1: Construct a causal game whose L_2-induced game is exactly G. We build an SCM with a plate over $t = 0, 1, 2, \ldots$ containing endogenous variables

$$S_t \in \mathcal{S}, \qquad X_{i,t} \in \mathcal{A}_i, \qquad Y_{i,t} \in \mathbb{R},$$

and exogenous variables

$$U_{i,t} \in \mathcal{A}_i \;\; (i \in N), \qquad U_t^S \text{ (transition noise).}$$

The structural equations are:

$$\begin{aligned}
S_0 &\sim \rho_0, & (11)\\
X_{i,t} &\leftarrow U_{i,t} \quad \text{(natural action mechanism)}, & (12)\\
S_{t+1} &\leftarrow f\big(S_t, X_t, U_t^S\big) \quad \text{such that } \Pr(S_{t+1} = s' \mid S_t = s, X_t = a) = P(s' \mid s, a), & (13)\\
Y_{i,t} &\leftarrow r_i(S_t, X_t) \qquad (i \neq 1), & (14)
\end{aligned}$$

and for player 1 we define a modified reward

$$Y_{1,t} \;\leftarrow\; r_1(S_t, X_t) \;+\; M\Big(\mathbf{1}\{U_{1,t} = X_{1,t}\} - \pi_1^\star(X_{1,t} \mid S_t)\Big), \tag{15}$$

where $M \in \mathbb{R}$ is a scalar parameter to be chosen.

Finally, set the exogenous distribution so that under the natural mechanism (12),

$$\Pr\big(U_{i,t} = a_i \mid S_t = s\big) = \pi_i^\star(a_i \mid s) \quad \text{for all } i \in N, \tag{16}$$

(and let U_t^S have whatever distribution is needed to implement (13)).

Define the discounted return of agent i as

$$R_i \;:=\; \mathbb{E}\Big[\sum_{t\geq 0} \gamma^t Y_{i,t}\Big].$$

Let Γ be the induced causal game with L_2 policies interpreted as interventions on the $X_{i,t}$ nodes (i.e., replacing (12) with a policy-dependent assignment).

Claim 1 (Correct L_2 semantics). Under L_2 actions (interventions on $X_{i,t}$), the induced sequential game is exactly the Markov game G.

Justification. Under any joint L_2 policy profile π, the interventions replace (12) by $X_{i,t} \leftarrow \pi_i(\cdot \mid \text{history})$ (or $X_{i,t} \leftarrow \pi_i(\cdot \mid S_t)$ for stationary policies). The transition equation (13) implements the Markov kernel $P(\cdot \mid s, a)$, and the base rewards are $r_i(S_t, X_t)$. Thus the distribution over trajectories and discounted returns coincides with that of G under π, except potentially for the extra term in (15), which we handle next.

[3] If G is infinite-horizon, we represent the repeated time-indexed variables using a plate model; all arguments below are per-time-step and therefore extend immediately.

[4] If multiple equilibria exist, fix an equilibrium-selection rule; the theorem holds for the selected equilibrium.

Claim 2 (The bonus term has zero interventional effect). For every time t and every intervention/policy profile (hence every realized $X_{1,t}$), the additional term in (15) has zero conditional expectation:

$$\mathbb{E}[\mathbf{1}\{U_{1,t} = X_{1,t}\} - \pi_1^\star(X_{1,t} \mid S_t) \mid S_t,\ \mathrm{do}(X_{1,t}),\ \mathrm{do}(X_{-1,t})] = 0.$$

Justification. Under an intervention, $X_{1,t}$ is set externally. Conditional on $S_t = s$ and $\mathrm{do}(X_{1,t} = a)$,

$$\mathbb{E}[\mathbf{1}\{U_{1,t} = X_{1,t}\} \mid S_t = s, \mathrm{do}(X_{1,t} = a)] = \Pr(U_{1,t} = a \mid S_t = s) = \pi_1^\star(a \mid s)$$

by (16), while $\pi_1^\star(X_{1,t} \mid S_t) = \pi_1^\star(a \mid s)$ deterministically. Hence the difference has zero expectation, so the interventional expected reward equals r_1. Therefore, the L_2-induced game (and its Nash equilibrium payoffs) are exactly those of G, and in particular the Nash equilibrium payoff in Γ equals μ_{NE}.

Step 2: Compute the L_1 payoff and shift it above/below μ_{NE}. Under L_1 behavior, actions follow the natural mechanism (12), so $X_{i,t} = U_{i,t}$. From (16), this means that conditional on $S_t = s$, $X_{i,t} \sim \pi_i^\star(\cdot \mid s)$, i.e., the L_1 trajectory distribution matches equilibrium play in G.

For $i \neq 1$, (14) implies $R_i^{L_1} = \mu_{\mathrm{NE},i}$. For $i = 1$, since $X_{1,t} = U_{1,t}$ almost surely under L_1, we have $\mathbf{1}\{U_{1,t} = X_{1,t}\} = 1$ almost surely, and thus

$$\mathbb{E}\big[\mathbf{1}\{U_{1,t} = X_{1,t}\} - \pi_1^\star(X_{1,t} \mid S_t) \mid S_t = s\big] = 1 - \sum_{a_1 \in \mathcal{A}_1} \pi_1^\star(a_1 \mid s)^2.$$

Let

$$\Delta \ := \ \mathbb{E}\Big[\sum_{t \geq 0} \gamma^t \Big(1 - \sum_{a_1} \pi_1^\star(a_1 \mid S_t)^2\Big)\Big].$$

Then the L_1 payoff for player 1 in Γ is

$$R_1^{L_1} \ = \ \mu_{\mathrm{NE},1} + M\Delta.$$

Note that $\Delta \geq 0$ always, and $\Delta > 0$ whenever $\pi_1^\star(\cdot \mid s)$ is non-degenerate on a set of states visited with positive discounted probability.[5]

Choose $M_+ > 0$ and $M_- < 0$ with $|M_\pm|$ large enough that $M_+\Delta > 0$ and $M_-\Delta < 0$. Let Γ_2 be the above construction with $M = M_+$ and let Γ_1 be the same construction with $M = M_-$. Then:

- By Claims 1–2, both Γ_1 and Γ_2 induce exactly the same Markov game G under L_2 actions, hence have the same Nash equilibrium payoff vector μ_{NE}.
- Under L_1 behavior, for all $i \neq 1$ we have $\mu_{1,i} = \mu_{\mathrm{NE},i} = \mu_{2,i}$, while for player 1,

$$\mu_{1,1} = \mu_{\mathrm{NE},1} + M_-\Delta \ < \ \mu_{\mathrm{NE},1} \ < \ \mu_{\mathrm{NE},1} + M_+\Delta = \mu_{2,1}.$$

Interpreting the inequalities componentwise in the standard Pareto order (i.e., $\mu_1 \leq \mu_{\mathrm{NE}} \leq \mu_2$ with at least one strict coordinate on each side), this establishes the claim. □

Example (One-state Markov game). Consider a two-player Markov game with a single state s, action sets $\mathcal{A}_1 = \mathcal{A}_2 = \{H, T\}$, and discount factor $\gamma \in (0, 1)$. The state is absorbing, i.e., $P(S_{t+1} = s \mid S_t = s, X_t) = 1$. Stage payoffs are given by the matching-pennies game: $r_1(H, H) = r_1(T, T) = 1$, $r_1(H, T) = r_1(T, H) = -1$, and $r_2 = -r_1$. This game admits a unique Nash equilibrium in stationary mixed strategies, where each player plays H and T with probability $1/2$, yielding equilibrium payoff vector $\mu_{\mathrm{NE}} = (0, 0)$.

We construct a causal game with endogenous variables $S_t, X_{1,t}, X_{2,t}, Y_{1,t}, Y_{2,t}$ and exogenous variables $U_{1,t}, U_{2,t}$. The natural (behavioral) action mechanism is given by $X_{i,t} \leftarrow U_{i,t}$ for $i \in \{1, 2\}$, with exogenous distribution $\Pr(U_{i,t} = H) =$

[5] If $\pi_1^\star$ happens to be pure on all states visited under itself, then $\Delta = 0$ and the construction yields $\mu_1 = \mu_{\mathrm{NE}} = \mu_2$ for player 1. In that case the strict inequalities in the theorem statement cannot hold componentwise for all players; the non-strict version $\mu_1 \leq \mu_{\mathrm{NE}} \leq \mu_2$ always holds.

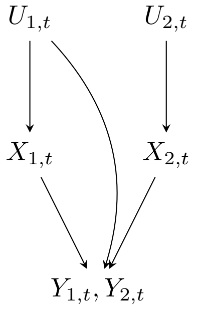


Figure 8. Per-time-step causal diagram for the matching-pennies causal game. The edge $U_{1,t} \to Y_{1,t}$ corresponds to the bonus term, which vanishes under intervention but shifts L_1 payoffs.

$\Pr(U_{i,t} = T) = \frac{1}{2}$. Rewards are defined as $Y_{2,t} \leftarrow r_2(X_{1,t}, X_{2,t})$ and $Y_{1,t} \leftarrow r_1(X_{1,t}, X_{2,t}) + M(\mathbf{1}\{U_{1,t} = X_{1,t}\} - \frac{1}{2})$, where $M \in \mathbb{R}$ is a scalar parameter.

Under interventions on $X_{i,t}$, the bonus term has zero expectation, since $\mathbb{E}[\mathbf{1}\{U_{1,t} = X_{1,t}\} - \frac{1}{2} \mid \mathrm{do}(X_{1,t})] = 0$. Consequently, the L_2-induced game coincides exactly with the original Markov game, and its Nash equilibrium payoff remains $\mu_{\mathrm{NE}} = (0, 0)$.

Under the natural mechanism, $X_{i,t} = U_{i,t}$ almost surely, so $\mathbb{E}[Y_{1,t}] = \frac{M}{2}$ and $\mathbb{E}[Y_{2,t}] = 0$. The discounted returns are therefore $R_1^{L_1} = \sum_{t \geq 0} \gamma^t \frac{M}{2} = \frac{M}{2(1-\gamma)}$ and $R_2^{L_1} = 0$. Choosing $M = -1$ yields a causal game Γ_1 with L_1 payoff $\mu_1 = (-\frac{1}{2(1-\gamma)}, 0)$, while choosing $M = +1$ yields a causal game Γ_2 with L_1 payoff $\mu_2 = (\frac{1}{2(1-\gamma)}, 0)$, and thus $\mu_1 < \mu_{\mathrm{NE}} < \mu_2$ while both causal games induce the same Nash equilibrium payoff under L_2 reasoning.

A.4. Proof of Theorem 3.5 (Dominance of Causal Strategies)

Proof. Let Γ be a (sequential) causal game and let L_Γ denote its *layer–selection game*, in which each agent chooses a subset of admissible causal policies (equivalently, a subset of PCH layers). Assume that L_Γ admits a pure–strategy Nash equilibrium $A^* = (A_1^*, \ldots, A_n^*)$, where A_i^* denotes the policy class selected by agent i. Let

$$\mu^* \;=\; \mathrm{NE}\big(\Gamma(A^*)\big)$$

be the Nash–equilibrium payoff vector of the causal game Γ when each agent is restricted to policies in A^*.

Fix an agent i and consider any alternative admissible policy class A_i' for that agent. Suppose, for the sake of contradiction, that deviating to A_i' strictly improves agent i's equilibrium payoff, i.e.,

$$\mathrm{NE}_i\big(\Gamma(A_i', A_{-i}^*)\big) \;>\; \mathrm{NE}_i\big(\Gamma(A^*)\big) \;=\; \mu_i^*.$$

By the definition of the layer–selection game L_Γ, the payoff to agent i from choosing a policy class is exactly the Nash–equilibrium payoff of the corresponding PCH–restricted causal game. Therefore, holding A_{-i}^* fixed, agent i would obtain a strictly higher payoff in L_Γ by switching from A_i^* to A_i'.

This contradicts the assumption that A^* is a Nash equilibrium of L_Γ. Hence, no such profitable unilateral deviation exists. It follows that for every agent i and every admissible alternative policy class A_i',

$$\mu_i^* \;\geq\; \mathrm{NE}_i\big(\Gamma(A_i', A_{-i}^*)\big).$$

This establishes the claim. □

B. Extensive Form Games and Causal Games

Below we propose an algorithm to convert an extensive form game to causal game. This also shows that for every Extensive Form Game, a causal Game exists that entails the extensive Form game under L_2 actions.

Algorithm 3 `EFG to CG`

```
Input: Extensive-form game $G = (N, H, Z, A, P, \{\mathcal{I}_i\}_{i \in N \cup \{c\}}, \pi, \{u_i\}_{i \in N})$
Output: A Causal Game $\Gamma$ induced by an SCM $M = \langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}) \rangle$
Initialize an empty SCM with endogenous node set $\mathbf{V} \leftarrow \emptyset$ and edge set $E \leftarrow \emptyset$.
Initialize exogenous node set $\mathbf{U} \leftarrow \emptyset$.
For each player $i \in N$ and each information set $I \in \mathcal{I}_i$, create a decision node $X_I$ with domain $A(I)$ and set $\mathbf{X}_i \leftarrow \mathbf{X}_i \cup \{X_I\}$, $\mathbf{V} \leftarrow \mathbf{V} \cup \{X_I\}$. {$A(I) := A(h)$ for any $h \in I$ (well-defined in an EFG)}
Let $\mathcal{I}_c$ denote the collection of chance information sets. For each $J \in \mathcal{I}_c$, create an endogenous node $V_J$ with domain $A(J)$ and add $V_J$ to $\mathbf{V}$.
Create an exogenous node $U_J$ for each $J \in \mathcal{I}_c$, add $U_J$ to $\mathbf{U}$, and set its distribution to match $\pi(\cdot \mid h)$ for any $h \in J$.
Define the structural assignment for chance as $V_J \leftarrow U_J$.
For each player information set $I \in \mathcal{I}_i$, compute the set $\mathrm{Obs}(I)$ of prior information sets whose realized actions are observed by player $i$ at $I$.
For each $K \in \mathrm{Obs}(I)$:
    • if $K \in \mathcal{I}_j$ for some player $j$, add edge $X_K \to X_I$;
    • if $K \in \mathcal{I}_c$, add edge $V_K \to X_I$.
Create a terminal outcome node $T$ with domain $Z$ and add $T$ to $\mathbf{V}$.
For every decision or chance node $X \in \{X_I\}_{I \in \cup_i \mathcal{I}_i} \cup \{V_J\}_{J \in \mathcal{I}_c}$, add an edge $X \to T$. {Equivalently, add only those nodes that can affect reachability of terminal histories.}
Define $T$ deterministically: for each assignment to $\mathrm{Pa}(T)$, forward-simulate play in $G$ to obtain the unique terminal history $z \in Z$, and set $T \leftarrow z$.
For each player $i \in N$, create a utility node $Y_i$ with parent $T$ and define
    $$Y_i \leftarrow u_i(T),$$
i.e., $Y_i(T = z) := u_i(z)$ for all $z \in Z$. Add $Y_i$ to $\mathbf{V}$.
Return: CG $\Gamma$ induced by SCM $M$
```

C. Experiments

C.1. Parameters for the Kuhn Poker

In Kuhn Poker, the body language correctly signals the card to the intuition with probability 0.8 and rest of the time, it signals the other card. For example, if the opponent has Jack and the player has a Queen, then the natural intuition of the player is to bet 80% of the time and fold the remaining times. The agents are dealt cards and the game proceeds in a round robin fashion.

C.2. Computation for the Iterated Causal Prisoner's Dilemma

The payoff of playing defect against intuition in M_1 can be easy calculated as follows:

$$E[\mathbf{Y} \mid do(X_1 = 1)] = 0.16 \cdot (-3, 0) + 0.24 \cdot (-8, 0) = (-2.4, 0) \tag{17}$$

When P_1 is playing Tit-for-Tat, it simply observe player 2 playing $X_2 = 1$ 60% time and $X_2 = 0$ the remaining 40%. Since, each timesteps assignment of U_1, U_2 are independent, this is equivalent to playing $X_1 = 1$ 60% of the time as well. This gives the payoff

$$E[\mathbf{Y} \mid do(X_1 = 1)] \cdot 0.6 + E[\mathbf{Y} \mid do(X_1 = 0)] \cdot 0.4 = (-3.56, 0.08) \tag{18}$$

The codes are available at https://anonymous.4open.science/r/CGT-ICML26/.