1 CORRECTNESS PROOF FOR LEARNCID

1.1 Setup and Assumptions

Our goal is to recover the causal structure and the Conditional Probability Tables (CPTs) of the variables describing the environment in which the agents operate. This environment consists of both observable variables and hidden latent variables. We define a set of interventions modeling distribution shifts, and to learn the causal graph, we query an optimal policy oracle associated with one agent to get the optimal policy for that agent under the specified distribution shift. Observe that this setup is unsuitable for traditional causal discovery algorithms like PC [8] and FCI [7] because we do not have access to the joint probability distribution of the variables or any sample data.

To model the causal relationships in the environment, we use Causal Influence Diagrams (CIDs) [1, 2]. Similar to Influence Diagrams [3], CIDs are commonly used to reason about decision-making tasks. CIDs further assume that the graph encodes the causal relationships between the nodes. We denote the set of parents of a node X as Pa_X, the set of children as Ch_X, the set of ancestors as Anc_X, the set of descendants as $Desc_X$ and instantiations of random variables in lower-case.

Definition 1 (Causal influence diagram [1, 2]). A *causal influence diagram* (CID) is a Causal Bayesian Network $M = (G = \{V, E\}, P)$, where P is a joint probability distribution compatible with the conditional independences encoded in G. The variables in V are partitioned into decision, utility, and chance variables, $V = (D, U, C)$. Each utility node U_i is associated with a real function f_i of its parents $f_i : \mathrm{Im}(Pa_{U_i}) \to \mathbb{R}$.

Each agent may correspond to a different set of decision nodes and have access to a distinct subset of observable variables. A variable observed by one agent may be latent for another. Additionally, considering a situation where an agent takes more than one decision, the set of variables that he observes when it takes one decision can differ from the one that he observes when taking another decision.

Following the work of [6], we represent distribution shifts as mixtures of local interventions. Given a random variable X with $x_1, \ldots, x_n$ as possible observable values, a local intervention on X is a function $\sigma : x_i \mapsto f(x_i)$ that maps each observable value x_i to a new observable value $f(x_i)$. In other words, local interventions deterministically reassign a random variable's outcomes independently of other variables.

Definition 2 (Local intervention [6]). *Local intervention* σ on X involves applying a map to the states of X that is not conditional on any other endogenous variables, $x \mapsto f(x)$. We use the notation $\sigma = do(X = f(x))$ (variable X is assigned the state $f(x)$). Formally, this is a soft intervention on X that transforms the conditional probability distribution as,

$$P(x \mid \mathbf{pa}_X; \sigma) \coloneqq \sum_{x' : f(x') = x} P(x' \mid \mathbf{pa}_X) \tag{1}$$

In general, a local intervention has limited capacity to model distribution shifts. For instance, it cannot model the shift from a coin that always lands on heads to a fair coin because a local intervention must deterministically map the observable value 'head' to another observable value. Therefore, we now report the concept of a mixture of local interventions [6]. This mixture is a convex combination $\sigma^* = \sum_i p_i \sigma_i$ of local interventions σ_i, where each coefficient p_i represents the probability that σ_i is used to map the observable value for X.

Definition 3 (Mixtures of interventions [6]). A *mixed intervention* $\sigma^* = \sum_i p_i \sigma_i$ for $\sum_i p_i = 1$ performs intervention σ_i with probability p_i. Formally, $P(x \mid \sigma^*) = \sum_i p_i P(x \mid \sigma_i)$.

We use optimal policy oracles to formalize the agent's understanding of optimal behavior under distribution shifts. Let D be a decision variable with observable values $d \in \mathrm{Im}(D)$, given a set of interventions Σ, an optimal policy oracle is a map $\Pi^*_\Sigma : \sigma \mapsto \pi_\sigma(d \mid pa_D)$ for $\sigma \in \Sigma$, where $\pi_\sigma(d \mid pa_D)$ is the optimal policy under the distribution shift induced by the intervention σ.

Definition 4 (Policy oracle). A policy oracle for a set of interventions Σ is a map $\Pi^*_\Sigma : \sigma \mapsto \pi_\sigma(d \mid pa_D)\ \forall\ \sigma \in \Sigma$, where $\pi_\sigma(d \mid pa_D)$ is an optimal policy under the intervention σ.

In our work, we rely on Algorithm 1 from [6], which takes as input a utility function U, an optimal policy oracle, an intervention $\sigma \in \Sigma$, and a parameter N that controls the number of samples. For any local intervention $\sigma \in \sigma_Y$, let d be the deterministic optimal decision under the shift induced by σ. By Assumption 9, there exists a hard intervention σ' such that d is no longer optimal. Let d_2 be the deterministic optimal decision under σ'. Considering the mixture $\sigma(q) \coloneqq q\sigma + (1-q)\sigma'$, there exist a value q_{crit} for q such that d_2 and another decision d_1 are both optimal. The algorithm returns q_{crit}, d_1, and d_2.

Now, we list and motivate our assumptions:

Assumption 1. Given the CID $M = (G = \{V, E\}, P)$ with $V = (D, U, C)$, the set of nodes and the partition (D, U, C) is known.

The set of nodes together with the node partition (D, U, C) is known, therefore we know all variables in the system and the type of each node (decision, utility, or chance).

Assumption 2. The CID is faithful [9] and sufficient [5].

Faithfulness implies that every conditional independence encoded in the graph G also holds in the joint probability P. A set of variables in a causal model is sufficient when it includes all common causes.

Assumption 3. The CID contains exactly one decision node D and one utility node U.

Despite assumption 3, this algorithm can be applied to multi-decision CIDs. It is possible to find further details in the main paper.

Assumption 4. The Markov blanket of decision node D is known. We also know all the edges between these nodes. The CPTs of chance nodes that are children of D are known.

Motivations for Assumption 4 can be found in Section 2.

Assumption 5. All chance nodes are ancestors of U.

In the presence of chance nodes that are neither ancestors of D or U, these nodes do not have any influence on the decision task. LearnCID would simply not process those nodes and the related causal structure or CPT would not be recovered.

Assumption 6. The utility function f associated with the utility node U is fully specified.

The utility function's functional form is known, which tells us all the variables involved in calculating the utility. These variables appear in the causal graph as parents of the utility node.

Assumption 7. We have access to a set Σ of all possible mixtures of local interventions, along with the optimal policy oracle Π^*_Σ for decision node D.

We consider single-decision, single-utility CIDs. To simplify the proof we assume D is a parent of U. In particular, this plays a role in the proof of Lemma 1.

Assumption 8. D is a parent of U.

Assumption 9. There exists no $d^* \in Im(D)$ such that $d^* \in \arg\max_d U(d, x)\ \forall\, x \in Im(Pa_U \setminus \{D\})$.

When Assumption 6 holds, it is possible to verify if the CID we are considering satisfies Assumption 9 by computing the utility for different instantiations of variables associated with the parents of U. Observe that under Assumptions 3, 8 and 9, there must be at least one chance node that is a parent of U, otherwise the utility function would only depend on the decision and therefore there would exist at least one optimal decision that would violate Assumption 9. We provide a discussion the Assumption 9 and domain dependence in Section 3.

1.2 Proof

Lemma 1. *Under Assumptions2,3,6 and 8, given a CID $M = (G, P)$, for any given local intervention σ there is a single deterministic optimal policy for almost all P, U.*

Proof. When $Desc_D \cap Anc_U = \emptyset$ we can apply Lemma 3 in [6]. Assume $Desc_D \cap Anc_U \neq \emptyset$ and $Ch_D \cap C \neq C$, let $\mathbf{Z} = Anc_U \setminus Pa_D$, and $\mathbf{X} = Pa_U$ Assume $\exists\, d_1, d_2$ s.t. both are optimal decisions and:

$$\mathbb{E}[u|Pa_D, do(D = d_1); \sigma] = \mathbb{E}[u|Pa_D, do(D = d_2); \sigma] \tag{2}$$

Equivalently:

$$\sum_z U(d_1, x)P(z, pa_D|do(D = d_1); \sigma) = \sum_z U(d_2, x)P(z, pa_D|do(D = d_2); \sigma) \tag{3}$$

$$\sum_z U(d_1, x)P(z, pa_D|do(D = d_1); \sigma) - U(d_2, x)P(z, pa_D|do(D = d_2); \sigma) = 0 \tag{4}$$

Let us define $S_1 \coloneqq \mathbf{Z} \setminus Ch_D$ and $\overline{Pa}_{C_i} \coloneqq Pa_{C_i} \setminus \{D\}$. We factorize the joint distribution:

$$P(z, pa_D|do(D = d_1); \sigma) \tag{5}$$

as:

$$\prod_{Z_i \in S_1} P(Z_i|pa_{Z_i}; \sigma) \prod_{C_i \in Ch_D} P(C_i|\overline{pa}_{C_i}, do(D = d_1); \sigma) \tag{6}$$

Same for d_2. Now, we can rewrite Equation 4 as:

$$\begin{aligned}\sum_z \prod_{Z_i \in S_1} P(Z_i|pa_{Z_i}; \sigma)\Big[U(d_1, x) \prod_{C_i \in Ch_D} P(C_i|\overline{pa}_{C_i}, do(D = d_1); \sigma) \\ - U(d_2, x) \prod_{C_i \in Ch_D} P(C_i|\overline{pa}_{C_i}, do(D = d_2); \sigma)\Big] = 0\end{aligned} \tag{7}$$

Since for all $C_i \in Ch_D$, $C_i \perp\!\!\!\perp D | \overline{Pa}_{C_i}$ in $G_{\underline{D}}$, then according to rule 2 of do-calculus [5] we can rewrite the equation as:

$$\sum_z \prod_{Z_i \in S_1} P(Z_i | pa_{Z_i}; \sigma) \Big[U(d_1, x) \prod_{C_i \in Ch_D} P(C_i | \overline{pa}_{C_i}, D = d_1; \sigma) - U(d_2, x) \prod_{C_i \in Ch_D} P(C_i | \overline{pa}_{C_i}, D = d_2; \sigma) \Big] = 0 \tag{8}$$

We can define $a(d, z) \coloneqq U(d, x) \prod_{C_i \in Ch_D} P(C_i | \overline{pa}_{C_i}, do(D = d); \sigma)$.
Observe that:

$$\sum_z \prod_{Z_i \in S_1} P(Z_i | pa_{Z_i}; \sigma) \big[a(d_1, z) - a(d_2, z) \big] = 0 \tag{9}$$

Is a polynomial equation with variables $\forall\, i.\ P(Z_i | pa_{Z_i}; \sigma)$. Also notice that if we rewrite each of these variables using Definition 2 (Local intervention) $P(Z_i | pa_{Z_i}; \sigma) = \sum_{z'_i : f(z'_i) = z_i} P(c'_i | pa_{Z_i})$ the equation is still a polynomial equation with all the parameters of the CPTs, excluding those related to the children of D, as variables. If the polynomial is not trivial ($\exists\, d, z.\ a(d_1, z) - a(d_2, z) \neq 0$) then the Lebesgue measure of the solution of this equation is zero [4], and since the set of parameters that allow for multiple optimal solutions has measure zero along at least one dimension, it follows that the whole set has measure zero.

Now, let us consider when the polynomial is trivial. Then for all d, z we have:

$$a(d_1, z) - a(d_2, z) = 0 \tag{10}$$

$$U(d_1, x) \prod_{C_i \in Ch_D} P(C_i | \overline{pa}_{C_i}, D = d_1; \sigma) - U(d_2, x) \prod_{C_i \in Ch_D} P(C_i | \overline{pa}_{C_i}, D = d_2; \sigma) = 0 \tag{11}$$

Again, this is a finite number of polynomial equations with some of the network parameters as variables and its coefficients are not trivial because $d_1 \neq d_2$ and because of Assumption 8. Therefore, it is satisfied only on a set of Lebesgue measure zero [4].

Now assume $Ch_D \cap C = C$, then the factorization of Equation 6 simplifies to:

$$\prod_{C_i \in Ch_D} P(C_i | \overline{pa}_{C_i}, do(D = d_1); \sigma) \tag{12}$$

And, by applying rule 2 of do-calculus [5] as before we can rewrite Equation 4 as:

$$U(d_1, x) \prod_{C_i \in Ch_D} P(C_i | \overline{pa}_{C_i}, D = d_1; \sigma) - U(d_2, x) \prod_{C_i \in Ch_D} P(C_i | \overline{pa}_{C_i}, D = d_2; \sigma) = 0 \tag{13}$$

That again is a polynomial equation in some of the network parameters and therefore the set of solutions has Lebesgue measure zero. Again, since the set of parameters that satisfies Equation 4 has measure zero along at least one dimension, the whole set has measure zero.

This implies that for almost all P, U and any given local intervention σ the optimal decision is unique.

□

Lemma 2. *Given a CID $M = (G, P)$, under Assumptions 2∼9, given an optimal policy oracle Π^*_Σ where Σ includes all mixtures of local interventions on C including masking inputs $Pa'_D \subset Pa_D$, then for any given $Pa'_D = pa'_D$ such that $Pa'_D \cap Pa_U = \emptyset$, we can identify:*

$$\sum_z P(C = c | do(D = d); \sigma) U(d, x) - P(C = c | do(D = d'); \sigma) U(d', x) \tag{14}$$

for $d, d' \in \mathrm{Im}(D)$ where $d \neq d'$ and $\mathbf{Z} = C \setminus Pa'_D$.

Proof. By Lemma 1, for almost all P, U there exist only one decision $d_1 = \arg\max_d \mathbb{E}[u | do(D = d), pa'_D; \sigma]$ following the shift σ. Let us call this decision d_1, it can be identified using the optimal policy oracle $\Pi^*_\Sigma(\sigma)$.

Thanks to Assumption 9 we know that for every decision d in the context $Pa'_D \subset Pa_D$, there exist at least one instance $c = (c_1, \ldots, c_N)$ of C where $d \neq \arg\max_{d'} U(d', x)$. Mind that we can set the values for X as $Pa'_D \cap Pa_U = \emptyset$. So let x' be the instantiation of $Pa_U \setminus \{D\}$ that satisfies $d_1 \neq \arg\max_d U(d, x')$, and σ' be an hard intervention that sets X to x', then there exist $d_2 = \arg\max_d U(d, x')$, with $d_2 \neq d_1$. Note that under an hard intervention like σ' we have $\mathbb{E}[u | do(D = d), pa'_D; \sigma] = U(d, x')$ where x' are the values that the variables $Pa_U \setminus \{D\}$ take after the intervention. We can pick σ' such that it sets Pa'_D to be the same as in observation.

	$(Anc_D \cup Desc_D)^C$	Anc_D	$Desc_D$
$Anc_U(G_{\underline{D}})$	(1)	(1)	(3)
$Anc_U(G_{\underline{D}})^C$	$\emptyset$	(2)	$\emptyset$

Table 1: A partition of the CID's chance nodes C that assigns them to cases in Theorem 1. Cells marked with (1) correspond to nodes that can be identified by case 1 of Theorem 1 in [6]. Cells with (2) correspond to case 2. Cells with $\emptyset$ indicate that either no nodes fall within that intersection or that those nodes can be pruned since we are not able to learn the structure or the parameters for those. The cell marked with (3) corresponds to nodes that exist only in the mediated case and were therefore not considered by the previous paper.

For $q \in [0,1]$ consider the joint distribution over C under the parametrized family of mixed local interventions $\tilde{\sigma}(q) = q\sigma + (1-q)\sigma'$:

$$P(C=c|do(D=d);\tilde{\sigma}(q)) = qP(C=c|do(D=d);\sigma) + (1-q)P(C=c|do(D=d);\sigma') \tag{15}$$

From Assumption 9 it follows that $Z \coloneqq C \setminus Pa_D \neq \emptyset$. Otherwise the Lemma's statement would be trivially false. We can write the expected utility as:

$$\mathbb{E}[U|pa_d, do(D=d);\tilde{\sigma}(q))] = \sum_z P(Z=z|pa_D, do(D=d);\tilde{\sigma}(q))U(d,x) \tag{16}$$

$$= \sum_z \frac{P(C=c|do(D=d);\tilde{\sigma}(q))}{P(Pa_D = pa_D|do(D=d);\tilde{\sigma}(q))}U(d,x) \tag{17}$$

$$= \frac{1}{P(Pa_D = pa_D;\tilde{\sigma}(q))}\sum_z qP(C=c|do(D=d);\sigma)U(d,x) + (1-q)P(C=c|do(D=d);\sigma')U(d,x') \tag{18}$$

Where in Equation 18 $P(Pa_D = pa_D|do(D=d);\tilde{\sigma}(q)) = P(Pa_D = pa_D;\tilde{\sigma}(q))$ according to Rule 3 of do-calculus since $D \perp\!\!\!\perp Pa_D$ in $G_{\overline{D}}$ [5]. Note that d_1 is the optimal decision for $q=1$, but that is not the case for $q=0$. Therefore there exists q_{crit} such that for all $q < q_{crit}$ $d_2 \coloneqq \Pi^*_{\Sigma}(\tilde{\sigma}(q))$ is a decision in the set $\{d|d = \arg\max_d U(d,x')\}$, and for $q \geq q_{crit}$ the optimal decision is not in this set. Let $d_3 \notin \{d|d = \arg\max_d U(d,x')\}$. Consider the following equation:

$$\mathbb{E}[U|pa_D, do(D=d_2);\tilde{\sigma}(q_{crit})] - E[U|pa_D, do(D=d_3);\tilde{\sigma}(q_{crit})] = 0 \tag{19}$$

$$\begin{aligned}\iff \quad q_{crit}\left[\sum_z P(C=c|do(D=d_2);\sigma)U(d_2,x) - P(C=c|do(D=d_3);\sigma)U(d_3,x)\right] + \\ +(1-q_{crit})[U(d_2,x') - U(d_3,x')] = 0\end{aligned} \tag{20}$$

$$\iff \quad q_{crit} = \left(1 - \frac{\sum_z P(C=c|do(D=d_2);\sigma)U(d_2,x) - P(C=c|do(D=d_3);\sigma)U(d_3,x)}{U(d_2,x') - U(d_3,x')}\right)^{-1} \tag{21}$$

Therefore, since the functional relationship between U and its parents is known, if we find q_{crit} we can identify:

$$\sum_z P(C=c|do(D=d_2);\sigma)U(d_2,x) - P(C=c|do(D=d_3);\sigma)U(d_2,x) \tag{22}$$

□

Let $G_{\underline{D}}$ be G without the edges leaving D, the mediated case allows for $Desc_D \cap Anc_U(G_{\underline{D}}) \neq \emptyset$. Consider the partition of C proposed in Table 1, the proof of Theorem 1 in [6] can still be used with minor changes in the mediated case for some of the nodes, but a new case needs to be introduced for $Anc_U(G_{\underline{D}}) \cap Desc_D$.

Theorem 1. *For almost all CIDs $M = (G = \{V,E\}, P)$ satisfying Assumptions 2∼9, and where for each $C_i \in C$, we know a set $\widehat{Pa}_{C_i} \subset Pa_{C_i}$ and a set of nodes $V_{known} \subset V$ where $C_i \in V_{known} \iff \widehat{Pa}_{C_i} = Pa_{C_i}$. It is possible to identify G and the joint distribution P over all the ancestors of the utility node Anc_U except Ch_D given $\{\pi^*_\sigma(d|pa_D)\}_{\sigma\in\Sigma}$ where $\pi^*_\sigma(d|pa_D)$ is an optimal policy in the domain σ and Σ is the set of all mixtures of local interventions.*

Proof. Let $G_{\underline{D}}$ be G without the edges leaving D. Following the CID's chance node partition described in Table 1, we consider three cases:

- [Case 3, $Anc_U(G_{\underline{D}}) \cap Desc_D$]. For the third case, we provide a constructive proof for nodes in $Anc_U(G_{\underline{D}}) \cap Desc_D$. We establish this proof by strong induction. Consider a directed path $C_k \to \cdots \to C_1$ where $C_1 \in Pa_U, \forall\ i.\ C_i \in Desc_D \setminus Ch_D$. Assume we know Pa_{C_i} and $P(C_k|Pa_{C_i})$ for all $i = 1, \ldots, k-1$, we want to learn Pa_{C_k} and $P(C_k|Pa_{C_k})$. For each of the nodes Y_i in $Y \coloneqq C \setminus \{C_1, \ldots, C_k\}$ we define the following hard interventions $\sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa) \coloneqq do(Y_1 = y_1, \ldots, Y_i = \kappa, \ldots, Y_{|Y|} = y_n, C_k = f(c_k))$ where y is an instantiation for $Y \setminus \{Y_i\}$ and κ one for Y_i. Here $f(C_k)$ is the following local intervention on C_k:

$$f(C_k) = \begin{cases} c'_k, & C_k = c'_k \\ c''_k, & \text{otherwise} \end{cases} \tag{23}$$

  We also mask all inputs to the policy: $Pa'_D = \emptyset$. Assume $C_k \notin Ch_D$, by Lemma 2 we can identify the following query:

$$\begin{aligned} &\sum_c P(C = c|do(D = d); \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa))U(d, x)- \\ &-P(C = c|do(D = d'); \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa))U(d', x) = \end{aligned} \tag{24}$$

$$\begin{aligned} \sum_{c_k} \cdots \sum_{c_1} \prod_{j=1,\ldots,k} &P(C_j = c_j|pa_{C_j}, do(D = d); \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa))U(d, x)- \\ &- \prod_{j=1,\ldots,k} (C_j = c_j|pa_{C_j}, do(D = d); \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa))U(d', x) \end{aligned} \tag{25}$$

  According to Rule 3 of do-calculus [5], since $D \perp\!\!\!\perp C_1, \ldots, C_k|Y$ in $G_{\overline{Y}}$ the expression in Equation 25 is equal to:

$$= \sum_{c_k} \cdots \sum_{c_1} \prod_{j=1,\ldots,k} P(C_j = c_j|pa_{C_j}; \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa))[U(d, x) - U(d', x)] \tag{26}$$

$$= \sum_{c_k} P(C_k = c_k|pa_{C_k}; \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa))\beta(c_k) \tag{27}$$

$$= \sum_{c_k} P(C_k = c_k|pa_{C_k})\beta(c_k) \tag{28}$$

  where:

$$\beta(c_k) \coloneqq \sum_{c_{k-1}} \cdots \sum_{c_1} \prod_{j=1,\ldots,k-1} P(C_j = c_j|pa_{C_j}; \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa))[U(d, x) - U(d', x)] \tag{29}$$

  This result is analogous to the one for Case 1. In Equation 28, following the definition of the intervention $\sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa)$, we have $P(C_k = c'_k|pa_{C_k}; \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa)) = P(C_k = c_k|pa_{C_k})$ and $P(C_k = c''_k|pa_{C_k}; \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa)) = 1 - P(C_k = c'_k|pa_{C_k}; \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa)) = P(C_k = c_k|pa_{C_k})$. Therefore there is only one parameter to be identified. If $C_k \notin V_{known}$, we can repeat this procedure with a different leave-one-out intervention for each potential parent Y_i of C_k and different c_k. If for some configuration of κ_1, κ_2 and c_k we have $P(C_k = c'_k|pa_{C_k}; \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa_1)) \neq P(C_k = c'_k|pa_{C_k}; \sigma_{C_k}(Y \setminus Y_i = y, Y_i = \kappa_2))$ then $Y_i \in Pa_{C_k}$. We can exclude from this search all nodes in $\widehat{Pa}_{C_k}$, since we already know they are parents of C_k. If $C_k \in V_{known}$ we can skip this step since we already know Pa_{C_k}. Then, for each instantiations of the variables in Pa_{C_k} and each $c_k \in \mathrm{Im}(C_k)$ we repeat the procedure and recover all the parameters for C_k.

And now we describe the necessary modification to the cases covered in [6] (Case 1 and 2 in Table 1):

(1) $Anc_U(G_{\underline{D}}) \cap \left[(Anc_D \cup Desc_D)^C \sqcup Anc_D\right]$. The identification problem for nodes in this set is described in Theorem 1 Case 1 of [6]. The proof is based on strong induction on k for directed paths $C_k \to \cdots \to C_1$ where $C_1 \in Pa_U$, where for all $i = 1, \ldots, k$ we have $C_k \neq D$. The procedure to incorporate prior knowledge is the same as the one specified in the proof for Case 3.
(2) $Anc_U(G_{\underline{D}})^C \cap Anc_D$. This case corresponds to Theorem 1 Case 2 of [6]). The original proof considered strong induction on k for directed paths $C_k \to \cdots \to C_1$ where $C_1 \in Pa_D$. Again, the procedure to incorporate prior knowledge is the same as the one specified in the proof for Case 3.

□

2 NON-IDENTIFIABILITY OF Ch_D AND THEIR CPTS

Now we prove that, under the same assumptions made for the previous results, Ch_D cannot be uniquely determined in general. This motivates the LearnCID assumption requiring knowledge of the children of D, their CPTs, and their parents.

Theorem 2. *Let $M = (G = \{V, E\}, P)$ be a single decision/single utility CID, assume we know D, U, C, and $Pa(U), Pa(D)$. Let Σ be the set of all mixtures of local interventions, Π^*_Σ be the optimal policy oracle. Then in general Ch_D can not be uniquely determined for a set of parameters with a strictly positive Lebesgue measure.*

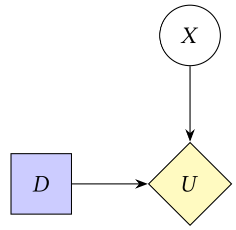


Figure 1: CID where $X \notin Ch_D$.

Proof. Consider the following counterexample:

For the CID described in Figure 1, $X \notin Ch(D)$. Let X, D be binary, and $P(X = 0) = 0.1$, $P(X = 1) = 0.9$. Let $U \coloneqq X$ AND D. In this case the optimal decision d^* is 1. The set of local interventions Σ contains $\sigma_1 \coloneqq do(X = 1)$, $\sigma_0 \coloneqq do(X = 0)$, and the parametrized familiy of mixtures $\sigma(q) \coloneqq q\sigma_1 + (1 - q)\sigma_0$ with $q \in [0, 1]$. Therefore, the policy oracle Π^*_Σ returns $d^* = 0$ for σ_0, $d^* = 1$ for σ_1, and for $\sigma(q)$ return $d^* = 0$ for $q < 0.5$, $d^* = 1$ for $q > 0.5$ and could return any policy for $q = 0.5$.

Now consider the CID described in Figure 2 and the CPT in Table 2, and let U and Σ be the same as in the previous example. Again, the optimal decision is $d^* = 1$ and again the policy oracle Π^*_Σ returns $d^* = 0$ for σ_0, $d^* = 1$ for σ_1, and for $\sigma(q)$ return $d^* = 0$ for $q < 0.5$, $d^* = 1$ for $q > 0.5$ and could return any policy for $q = 0.5$. So it is impossible to distinguish between the two models.

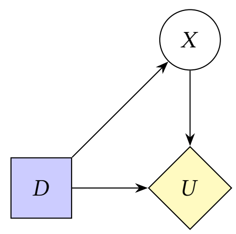


Figure 2: CID where $X \in Ch_D$.

D	X	$P(X \mid D)$
0	0	0.5
0	1	0.5
1	0	0
1	1	1

Table 2: Conditional probabilities of X given D for the CID where $D \in Ch_D$.

Also observe that this stays true if in the first CID we set $P(X = 1) \in (0.5, 1]$ and in the second CID we set $P(X = 1|D = 1) \in (0.5, 1]$ and $P(X = 0|D = 0) \in [0, P(X = 1|D = 1))$. So the set where Ch_D can not be uniquely determined has strictly positive Lebesgue measure. □

We also show that unlike for other chance nodes, the CPT of the children chance variables of the decision node can not be fully estimated in the general case. It might be possible to estimate only those parts of the CPT that correspond to decisions that are optimal for some distribution shift σ, but not for the others.

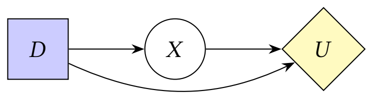
Corollary 3 (of Lemma 2). *Let $M = (G = \{V, E\}, P)$ be a single decision/single utility CID, assume we know G. Let Σ be the set of all mixtures of local interventions, Π^*_Σ be the optimal policy oracle. Then by using the identification result of Lemma 2, in general the CPTs for $C \cap Ch_D$ can not be uniquely determined for a set of parameters with a strictly positive Lebesgue measure.*

Proof. Consider the CID described in Figure 2 with the CPT described in Table 3. Assume $\mathrm{Im}(D) = \{0, 1, 2\}$ and that we don't know the CPT for X. The set of local interventions Σ contains $\sigma_1 \coloneqq do(X = 1)$, $\sigma_0 \coloneqq do(X = 0)$, and the parametrized familiy of mixtures $\sigma(q) \coloneqq q\sigma_1 + (1 - q)\sigma_0$ with $q \in [0, 1]$. Therefore, the policy oracle Π^*_Σ returns $d^* = 0$ for σ_0, $d^* = 1$ for σ_1, and for $\sigma(q)$ return $d^* = 0$ for $q < 0.5$, $d^* = 1$ for $q > 0.5$ and could return any policy for $q = 0.5$.

D	X	$P(X \mid D)$
0	0	0.5
0	1	0.5
1	0	0
1	1	1
2	0	0.4
2	1	0.6

Table 3: Conditional probabilities of X given D, this time $\mathrm{Im}(D) = \{0, 1, 2\}$.

D	X	$U(X\|D)$
0	0	0
0	1	1
1	0	0
1	1	2



D	X	$P(X\|D)$
0	0	0.5
0	1	0.5
1	0	0
1	1	1

D	X	$P'(X\|D)$
0	0	0
0	1	1
1	0	1
1	1	0

Figure 3: An example CID to show that in the mediated case domain dependence does not imply Assumption 9. Starting from the left, a specification of the utility function associated with node U, the example's CID, and the two CPTs for X before the distribution shift (P) and after the distribution shift (P').

Therefore, for all $\sigma \in \Sigma$ we have that the deterministic policy corresponding to $d = 3$ is never selected by the policy oracle $\Pi^*_\Sigma(\sigma)$. Since the identification result of Lemma 2 includes only probabilities $P(C|do(D = d'); \sigma)$ where d' is an optimal solution for some σ selected by the policy oracle, it means that in particular, we can't identify $P(X|do(D = 3)) = P(X|D = 3)$ or more in general the part of the CPT for chance variables that are children of D that correspond to solutions that are never optimal regardless of the intervention σ. □

3 DOMAIN DEPENDENCE IN MEDIATED TASKS

Previous results [6] show that for unmediated decision tasks domain dependence implies Assumption 9. Here we show that this implication does not hold in the mediated case. Moreover, we prove that Assumption 9 implies domain dependence in the mediated case and therefore that Assumption 9 is equivalent to domain dependence in the unmediated case, which is a subcase of the mediated one. We report the definition of Domain dependence.

Definition 5 (Domain dependence [6])**.** There exists $P(C = c)$ and $P'(C = c)$ compatible with M such that $\pi^* = \arg\max_\pi \mathbb{E}^\pi_P[U] \implies \pi^* \neq \arg\max_\pi \mathbb{E}^\pi_{P'}[U]$.

Consider the example described in Figure 3. Let $U(d, x) \coloneqq 2$ if $d = x = 1$, $U(d, x) \coloneqq 1$ if $d = 0$ and $x = 1$, and 0 otherwise. For the distribution P the only optimal policy corresponds to always choosing $D = 1$ because X will always correspond to 1 and therefore the expected utility is 2. But this policy is no longer optimal under P' because X will always be observed as 0 and the expected utility is 0 while for example a policy that always chooses $D = 0$ corresponds to an expected utility of 1. Therefore domain dependence holds, but at the same time, Assumption 9 does not hold because $d^* = 1 \in \arg\max_d U(d, x)$ for all $x \in \mathrm{Im}(X)$. Therefore, domain dependence $\not\Longrightarrow$ Assumption 9 in the general mediated case.

Now we prove that Assumption 9 implies domain dependence in the mediated case, and consequently is equivalent to domain dependence in the unmediated case.

Theorem 4. *Let $M = (G, P)$ be a mediated CID. Assumption 9 $\implies$ Domain dependence.*

Proof. Assume $\forall\ P'(C = c)$ compatible with M we have $\pi^* \in \arg\max_\pi \mathbb{E}[U] = \arg\max_\pi \mathbb{E}^\pi_{P'}[U|do(D = \pi(d|pa_D)), pa_D]$. Let $d \in \mathrm{Im}(D)$ be a decision s.t. $\pi^*(d|pa_D) > 0$. For Assumption 9 there exist a non-empty set $X_d \coloneqq \{x|d \notin \arg\max_{d'} U(d', x)\}$. Let $d^*, x^* \in \arg\max_{d', x \in X_d} U(d', x)$. We can write x^* as $(x^*_1, \dots, x^*_n)$ where $\{x^*_i\}^n_{i=1}$ are instantiations of the random variables $\{X_1, \dots, X_n\} = Pa_U \setminus \{D\}$ and $n \coloneqq |Pa_U \setminus \{D\}|$. Now, we want to define an alternative distribution P' compatible with M by updating the CPTs of the variables corresponding to the parents of U. For each $X_j \in Pa_u \setminus \{D\}$ let $pa^1_j, \dots, pa^{|Pa_{X_j}|}_j$ be instantiations of parents of X_j. Let x^i_j be an observable value for the variable X_j. We set $P(x^i_j|pa^1_j, \dots, pa^{|Pa_{X_j}|}_j) = \epsilon_{pa^1_j,\dots,pa^{|Pa_{X_j}|}_j, x^i_j}$ if $x^i_j \neq x^*_j$ and $P(x^i_j|pa^1_j, \dots, pa^{|Pa_{X_j}|}_j) = 1 - \sum_{l \neq i} \epsilon_{pa^1_j,\dots,pa^{|Pa_{X_j}|}_j, x^l_j}$ if $x^i_j = x^*_j$. We repeat this procedure for all combinations of x^i_j and $pa^1_j, \dots, pa^{|Pa_{X_j}|}_j$. We call the set of these epsilon parameters $\sum$. To preserve faithfulness we require the epsilon parameters to be pair-wise distinct. Observe that if we also require all epsilon $\epsilon \in \sum$ to be $0 <\ll 1$ and $0 < \sum_{\epsilon \in \sum} \epsilon \leq 1$ then we obtain valid CPT parameters for the CID. We repeat the CPT update for all variables $\{X_1, \ldots, X_n\}$.

Observe that assuming $\pi^* \in \arg\max \mathbb{E}^{\pi}_{P'}[U]$ for all P' compatible with M implies that $\forall\ \pi'$ and P' compatible with M we have:

$$\mathbb{E}^{\pi^*}_{P'}[U] - \mathbb{E}^{\pi'}_{P'}[U] \geq 0 \tag{30}$$

$$\iff \mathbb{E}_{P'}[U \mid do(D = \pi^*(d \mid pa_D) =,\ pa_D] - \mathbb{E}_{P'}[U \mid do(D = \pi'(d \mid pa_D),\ pa_D] \geq 0 \tag{31}$$

Let π' be a deterministic policy where d^* is always selected and $X \coloneqq Pa_U \setminus \{D\}$ with x instantiation of X.

$$\iff \mathbb{E}_{P'}[U \mid do(D = \pi^*(d \mid pa_D) =,\ pa_D] - \mathbb{E}_{P'}[U \mid do(D = d^*),\ pa_D] \geq 0 \tag{32}$$

$$\iff \sum_{d'} \pi^*(d' \mid pa_D) \sum_{C_i \in \mathbf{C}} \prod_j P(C_i = c_j \mid pa_{C_i}) U(d', x) - \sum_{C_i \in \mathbf{C}} \prod_j P(C_i = c_j \mid pa_{C_i}) U(d^*, x) \geq 0 \tag{33}$$

Now we compute the limit for all $\sum \ni \epsilon \longrightarrow 0$:

$$\lim_{\sum \ni \epsilon \to 0} \mathbb{E}^{\pi^*}_{P'}[U] - \mathbb{E}^{\pi'}_{P'}[U] = \tag{34}$$

$$= \sum_{d'} \pi^*(d' \mid pa_D) U(d', x^*) - U(d^*, x) \tag{35}$$

$$= \pi^*(d \mid pa_D) U(d, x^*) + \sum_{d' \neq d} \pi^*(d' \mid pa_D) U(d', x) - U(d^*, x^*) \leq \tag{36}$$

$$\leq \pi^*(d \mid pa_D) U(d, x^*) + U(d^*, x^*) \left[\sum_{d' \neq d} \pi^*(d' \mid pa_D) - 1 \right] \tag{37}$$

where in the last passage we used the fact that $d^* \in \arg\max_{d'} U(d', x^*)$.

$$= - \left[\sum_{d' \neq d} \pi^*(d' \mid pa_D) - 1 \right] U(d, x^*) + U(d^*, x^*) \left[\sum_{d' \neq d} \pi^*(d' \mid pa_D) - 1 \right] \tag{38}$$

$$= \pi^*(d \mid pa_D) \left(U(d^*, x^*) - U(d, x^*) \right) < 0 \tag{39}$$

The last expression is strictly negative because we assumed $\pi^*(d \mid pa_D) > 0$ and $x^* \in X_d$, therefore $d \notin \arg\max_{d'} U(d', x^*)$ while $d^* \in \arg\max_{d'} U(d', x^*)$. Since $\mathbb{E}^{\pi^*}_{P'}[U] - \mathbb{E}^{\pi'}_{P'}[U]$ is a polynomial in the parameters $\sum$, we can apply the theorem of permanence of sign and therefore $\exists \epsilon^* \in \sum$ s.t. the inequality 30 is false. Therefore $\exists P'$ compatible with M s.t. $\pi^* \notin \arg\max_{\pi} \mathbb{E}^{\pi}_{P'}$ $\lightning$. It follows that domain dependence holds. □

From Theorem 4 it directly follows that for unmediated decision tasks, which are a subcase of the family of mediated decision tasks, domain dependence is equivalent to Assumption 9. This provides us with a very straightforward way to verify domain dependence in these tasks.

Corollary 5. *Let $M = (G, P)$ be an unmediated CID. Assumption 9 is equivalent to Domain dependence.*

Proof. In the unmediated case the implication Domain dependence $\implies$ Assumption 9 is proven in [6]. Since the unmediated case is a subcase of the mediated case, from Theorem 4 it directly follows that Assumption 9 $\implies$ Domain dependence. Therefore the two statements are equivalent. □

REFERENCES

[1] Tom Everitt, Ryan Carey, Eric D. Langlois, Pedro A. Ortega, and Shane Legg. 2021. Agent Incentives: A Causal Perspective. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 13 (May 2021), 11487–11495. https://doi.org/10.1609/aaai.v35i13.17368

[2] David Heckerman. 1995. A Bayesian approach to learning causal networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (Montréal, Qué, Canada) *(UAI'95)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 285–295.

[3] Ronald A. Howard and James E. Matheson. 1984. Influence Diagrams. *Readings on the Principles and Applications of Decision Analysis, Vol. II* (1984).

[4] Masashi Okamoto. 1973. Distinctness of the Eigenvalues of a Quadratic form in a Multivariate Sample. *The Annals of Statistics* 1, 4 (1973), 763 – 765. https://doi.org/10.1214/aos/1176342472

[5] Judea Pearl. 2009. *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press, USA.

[6] Jonathan Richens and Tom Everitt. 2024. Robust agents learn causal world models. In *The Twelfth International Conference on Learning Representations.* https://openreview.net/forum?id=pOoKI3ouv1

[7] Peter Spirtes. 2001. An Anytime Algorithm for Causal Inference. In *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. R3)*, Thomas S. Richardson and Tommi S. Jaakkola (Eds.). PMLR, 278–285. https://proceedings.mlr.press/r3/spirtes01a.html Reissued by PMLR on 31 March 2021.

[8] Peter Spirtes and Clark Glymour. 1991. An algorithm for fast recovery of sparse causal graphs. *Social science computer review* 9, 1 (1991), 62–72.

[9] Peter Spirtes, Clark Glymour, N Scheines, et al. 1991. Causation, Prediction, and Search. (1991).