On the Granularity of Causal Effect Identifiability

Yizuo Chen[1] **Adnan Darwiche**[1]

[1]Computer Science Department, University of California, Los Angeles, USA

Abstract

The classical notion of causal effect identifiability is defined in terms of treatment and outcome variables. In this paper, we consider the identifiability of state-based causal effects: how an intervention on a particular *state* of treatment variables affects a particular *state* of outcome variables. We demonstrate that state-based causal effects may be identifiable even when variable-based causal effects may not. Moreover, we show that this separation occurs only when additional knowledge — such as context-specific independencies — is available. We further examine knowledge that constrains the states of variables, and show that such knowledge can improve both variable-based and state-based identifiability when combined with other knowledge such as context-specific independencies. We finally propose an approach for identifying causal effects under these additional constraints, and conduct empirical studies to further illustrate the separations between the two levels of identifiability.

1 INTRODUCTION

The causal effect is a distribution on some outcome variables that materializes as a result of an intervention on a set of treatment variables; e.g., the probability that a patient will recover if the doctor instructs them to take a drug. In general, a causal effect cannot be estimated based on observational data only, therefore requiring experimental studies; see, e.g., [Pearl and Mackenzie, 2018]. However, given a graph that encodes the causal relations among variables, a causal effect may in some cases be uniquely determined (i.e., estimable) using observational data only. In such cases, the causal effect is said to be *identifiable* and a formula may be derived to estimate it based on observational data.

Both complete and incomplete methods have been proposed in the past for testing the identifiability of causal effects based on causal graphs; see, e.g., [Pearl, 2009, Hernán and Robins, 2020] for some overviews. Some early methods for testing identifiability include the back-door criterion [Pearl, 1993, 2009] and front-door criterion [Pearl, 1995, 2009]. These methods are sound yet not complete and hence may fail to identify causal effects that are indeed identifiable. Complete methods for identifying causal effects include the do-calculus [Pearl, 2009], the identification algorithm in [Tian and Pearl, 2003, Huang and Valtorta, 2006], and the ID algorithm in [Shpitser and Pearl, 2006, 2008]. All these methods, however, are *variable-based* in that they test whether the causal effect is identifiable for *every* possible intervention on the treatment variables and *every* possible state of the outcome variables (for the same set of treatment and outcome variables). We will argue in this paper that variable-based identifiability is appropriate when the available information consists only of a causal graph and some observational data but can be too coarse-grained when additional information is available. In particular, given additional information, it is conceivable that the causal effect of *some* treatment on *some* outcome may be identifiable, while the causal effect of some other treatment on some other outcome may not be identifiable. This highlights the need for more refined (state-based) approaches to identifiability as we discuss and present later.

This distinction becomes particularly relevant in light of recent works that study the identifiability problem while assuming additional knowledge beyond causal graphs and observational data. This includes knowledge of functional dependencies [Chen and Darwiche, 2024], context-specific independencies (CSIs) [Tikka et al., 2019, Mokhtarian et al., 2022], and known parameters [Chen and Darwiche, 2025]. Such additional knowledge can be formulated as constraints, leading to the notion of *constrained-identifiability* [Chen and Darwiche, 2024, 2025]. The main use of such additional knowledge has been to show that it may render some (variable-based) causal effects identifiable which would otherwise be unidentifiable without the additional knowledge.

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:1181–1199.

We show in this paper that such knowledge has an additional value: it renders the distinction between variable-based causal effects and state-based causal effects meaningful and worthy of study as it highlights some new directions for improving causal effect identifiability.

Let $\mathbf{X}$ be the treatment variables and $\mathbf{Y}$ be the outcome variables. We start in Section 2 by formulating state-based identifiability which is concerned with whether the causal effect of a particular treatment ($\mathbf{X}$=$\mathbf{x}$) on a particular outcome ($\mathbf{Y}$=$\mathbf{y}$) can be uniquely determined based on available information. We then illustrate in Section 3 the distinction between state-based and variable-based identifiability given knowledge in the form of context-specific independencies (CSIs), which are effectively constraints on model parameters that cannot be captured by causal graphs. We follow in Section 4 by considering state constraints as an additional type of knowledge, showing that they are not useful on their own in the classical identifiability setup, but they may widen the gap between state-based and variable-based identifiability when coupled with other knowledge such as CSIs. We then present a novel method in Section 5 for identifying both variable-based and state-based causal effects under CSIs and state constraints, and follow by empirically evaluating its completeness and computationally efficiency in Section 6. We finally close with some concluding remarks in Section 7. All proofs are included in Appendix D.

2 TWO LEVELS OF CAUSAL EFFECT IDENTIFIABILITY

All variables are assumed to be discrete throughout this work. Let G denote a causal graph in the form of a DAG, $\mathbf{V}$ the set of observed variables in G, $\mathbf{X}$ the treatment variables, and $\mathbf{Y}$ the outcome variables. Variables outside $\mathbf{V}$ are called *hidden* and are denoted by $\mathbf{U}$. Let $\mathbf{x}$ and $\mathbf{y}$ denote some states (instantiations) of variables $\mathbf{X}$ and $\mathbf{Y}$, respectively.

A parameterization (model) of causal graph G assigns a conditional probability table (CPT) $\theta_{X|\mathbf{P}}$ to each variable X with parents $\mathbf{P}$ in G, representing the conditional distribution $\Pr(X|\mathbf{P})$. Such a model induces the distribution $\Pr(\mathbf{V},\mathbf{U}) = \prod_{X\in\mathbf{V}\cup\mathbf{U}} \theta_{X|\mathbf{P}}$ and the *observational distribution* $\Pr(\mathbf{V}) = \sum_{\mathbf{U}} \Pr(\mathbf{V},\mathbf{U})$. The probability of $\mathbf{y}$ under an intervention $\mathbf{x}$ is denoted $\Pr_{\mathbf{x}}(\mathbf{y})$ and computed using the *interventional distribution* $\Pr_{\mathbf{x}}(\mathbf{V},\mathbf{U})$ defined as follows; see [Pearl, 2009, Ch. 3]. We first remove the incoming edges of $\mathbf{X}$ in G and replace the CPT for each $X \in \mathbf{X}$ by a new CPT θ_X, where $\theta_X(x) = 1$ if $x \in \mathbf{x}$ and $\theta_X(x) = 0$ otherwise. The updated set of CPTs induces the interventional distribution $\Pr_{\mathbf{x}}(\mathbf{V},\mathbf{U})$. The classical identifiability problem checks whether any two models (parameterizations of G) that induce the same observational distribution $\Pr(\mathbf{V})$ also yield the same causal effect $\Pr_{\mathbf{x}}(\mathbf{y})$ for all $\mathbf{x}$ and all $\mathbf{y}$.

Definition 1 (Classical Identifiability)**.** *The causal effect of $\mathbf{X}$ on $\mathbf{Y}$, denoted $\Pr_{\mathbf{X}}(\mathbf{Y})$, is identifiable wrt $\langle G,\mathbf{V}\rangle$ if $\Pr^1_{\mathbf{x}}(\mathbf{y}) = \Pr^2_{\mathbf{x}}(\mathbf{y})$ for all instantiations $\mathbf{x},\mathbf{y}$ and any pair of models that induce $\Pr^1,\Pr^2$ such that $\Pr^1(\mathbf{V}) = \Pr^2(\mathbf{V})$.*[1]

Classical identifiability is variable-based in the following sense. If the causal effect $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable, then $\Pr_{\mathbf{x}}(\mathbf{y})$ is uniquely computable from the observational distribution $\Pr(\mathbf{V})$ for *every* possible treatment $\mathbf{x}$ and *every* possible outcome $\mathbf{y}$. Note that classical identifiability does not restrict the states of model variables.

We now introduce a more granular notion of identifiability which targets a particular treatment $\mathbf{x}$ and a particular outcome $\mathbf{y}$. For example, the probability that a major accident will occur if a particular speed limit is enforced, $\Pr_{\mathsf{speedLimit}=65}(\mathsf{accident}=\textit{major})$.

Definition 2 (State-Based Identifiability)**.** *The causal effect of $\mathbf{x}$ on $\mathbf{y}$, denoted $\Pr_{\mathbf{x}}(\mathbf{y})$, is identifiable wrt $\langle G,\mathbf{V}\rangle$ if $\Pr^1_{\mathbf{x}}(\mathbf{y}) = \Pr^2_{\mathbf{x}}(\mathbf{y})$ for any pair of models that induce $\Pr^1,\Pr^2$ such that $\Pr^1(\mathbf{V}) = \Pr^2(\mathbf{V})$.*

State-based identifiability refers to a particular treatment $\mathbf{x}$ and a particular outcome $\mathbf{y}$ which are states of the treatment and outcome variables. To illustrate the applicability of state-based identifiability, consider the following real-world scenarios. Suppose that speed limit signs (e.g., a maximum of 65 mph) have been installed on *some* streets in a district, and the policymakers are interested in the following question: how much would enforcing speed limits on *all* streets reduce the accident rate? This question requires comparing the state-based causal effect $\Pr_{\mathsf{SpeedLimit}=\textit{true}}(\mathsf{Accident}=\textit{true})$ with the observational probability $\Pr(\mathsf{Accident}=\textit{true})$. The policymakers are typically not interested in the accident rate under a policy that removes all speed limit signs, i.e., $\Pr_{\mathsf{SpeedLimit}=\textit{false}}(\mathsf{Accident}=\textit{true})$ — a quantity that would be computed in the variable-based causal effect $\Pr_{\mathsf{SpeedLimit}}(\mathsf{Accident})$ but is irrelevant in this scenario.

As another example, suppose a hospital is concerned with patients' recovery rates from a disease as well as the side effects of taking a drug. To decide whether it is beneficial to enforce *all* patients to take the drug, the hospital may be interested in the state-based causal effect $\Pr_{\mathsf{Drug}=\textit{true}}(\mathsf{Recover}=\textit{true}, \mathsf{SideEffects}=\textit{low})$, which is easier to obtain than the variable-based causal effect $\Pr_{\mathsf{Drug}}(\mathsf{Recover}, \mathsf{SideEffects})$ commonly considered for evaluating the drug's therapeutic effect. These two examples show that state-based causal effects become relevant when the focus is on the effect of an intervention relative to the current system (captured by the observational distribution), rather than on comparing the effects of distinct interventions as typically done in the classical setup.

[1]We assume that the observational distribution $\Pr(\mathbf{V})$ is strictly positive, i.e., $\Pr(\mathbf{V}) > 0$. This positivity assumption is common in existing identification methods such as IDENTIFY [Tian and Pearl, 2003] and the do-calculus [Pearl, 2009].

The distinction between variable-based and state-based identifiability is not interesting if the only information available consists of a causal graph G and an observational distribution $\Pr(\mathbf{V})$. In fact, these two notions are equivalent in the classical setup, which we formulate as the following proposition and include a proof in Appendix D.

Proposition 1. *The causal effect* $\Pr_{\mathbf{X}}(\mathbf{Y})$ *is identifiable iff* $\Pr_{\mathbf{x}}(\mathbf{y})$ *is identifiable for some states* $\mathbf{x}, \mathbf{y}$.

The distinction between the two notions materializes though if we have additional knowledge beyond G and $\Pr(\mathbf{V})$. We elaborate on this next.

3 THE IMPACT OF CSI CONSTRAINTS ON IDENTIFIABILITY

We next show how CSI constraints [Boutilier et al., 1996, Shen et al., 2020], studied for causal effect identifiability in [Tikka et al., 2019, Mokhtarian et al., 2022, Chen and Darwiche, 2025], draw a distinction between variable-based and state-based identifiability. We consider CSIs in the form of $(Y \perp\!\!\!\perp P_1 | \mathbf{p}_2, \mathbf{P}_3)$, where Y is a node in the causal graph and $\{P_1\}, \mathbf{P}_2, \mathbf{P}_3$ is a partition of its parents. This CSI constraint is interpreted as follows: Y is independent of P_1 under the specific state $\mathbf{p}_2$ of $\mathbf{P}_2$ and any states of $\mathbf{P}_3$. We call $\mathbf{P}_2$ the *context variables* and $\mathbf{p}_2$ the *context* of the CSI. For example, the CSI ($\mathsf{Light} \perp\!\!\!\perp \mathsf{People} \mid \mathsf{Outage}$=*yes*, Time) states that when a power outage occurs (context), the status of light becomes independent of the presence of people, regardless of the time of day. This independence, however, does not hold when the power outage does not occur. CSI constraints are effectively constraints on model parameters so they cannot be captured using the language of causal graphs. Constraints, such as CSIs, restrict the set of models considered when deciding identifiability. This has the impact of making some causal effects identifiable, which would be unidentifiable without the constraints. We assume that the constraints discussed in this work are known and provided as additional inputs to the identifiability problem. Our results are therefore independent of how these constraints are obtained, whether specified from domain knowledge or learned from data. To continue with this discussion, we need to recall the following definition of constrained identifiability (at both the variable-based and state-based levels).

Definition 3. *[Chen and Darwiche, 2025] Let $\mathcal{C}$ be a set of constraints. A causal effect* $\Pr_{\mathbf{x}}(\mathbf{y})$ *is* <u>*identifiable*</u> *wrt* $\langle G, \mathbf{V}, \mathcal{C}\rangle$ *if* $\Pr^1_{\mathbf{x}}(\mathbf{y}) = \Pr^2_{\mathbf{x}}(\mathbf{y})$ *for distributions* $\Pr^1, \Pr^2$ *that are induced by G, satisfy constraints $\mathcal{C}$ and* $\Pr^1(\mathbf{V}) = \Pr^2(\mathbf{V})$. *Moreover, the causal effect* $\Pr_{\mathbf{X}}(\mathbf{Y})$ *is* <u>*identifiable*</u> *wrt* $\langle G, \mathbf{V}, \mathcal{C}\rangle$ *if* $\Pr_{\mathbf{x}}(\mathbf{y})$ *is identifiable for all* $\overline{\mathbf{x}, \mathbf{y}}$.

Our first result states that the (state-based) identifiability of $\Pr_{\mathbf{x}}(\mathbf{y})$ and the (variable-based) identifiability of $\Pr_{\mathbf{X}}(\mathbf{Y})$ are equivalent if none of the states $\mathbf{x}, \mathbf{y}$ appear in the contexts of CSIs. For example, if ($\mathsf{Light} \perp\!\!\!\perp \mathsf{People} \mid \mathsf{Outage}$=*yes*, Time) is the only CSI constraint, then $\Pr_{\mathsf{Outage}=no}(\mathsf{Light}=off)$ is identifiable iff $\Pr_{\mathsf{Outage}}(\mathsf{Light})$ is identifiable. The result is stated in the following proposition.

Proposition 2. *Consider Definition 3 of constrained identifiability and let $\mathcal{C}$ be a set of CSIs. If none of the states* $\mathbf{x}, \mathbf{y}$ *appear in the contexts of $\mathcal{C}$, then* $\Pr_{\mathbf{x}}(\mathbf{y})$ *is identifiable wrt* $\langle G, \mathbf{V}, \mathcal{C}\rangle$ *iff* $\Pr_{\mathbf{X}}(\mathbf{Y})$ *is identifiable wrt* $\langle G, \mathbf{V}, \mathcal{C}\rangle$.

It immediately follows that the (variable-based) identifiability of $\Pr_{\mathbf{X}}(\mathbf{Y})$ can be reduced to the (state-based) identifiability of $\Pr_{\mathbf{x}}(\mathbf{y})$ for states $\mathbf{x}, \mathbf{y}$ that do not appear in CSI contexts. Hence, if we have an algorithm for testing state-based identifiability under CSIs, we can use it to test variable-based identifiability under CSIs by choosing states $\mathbf{x}, \mathbf{y}$ that do not appear in CSI contexts. This can always be done as we are only considering CSI constraints and these do not constrain variable states.

However, in general, state-based and variable-based identifiability are not equivalent. We demonstrate the separation between the two using the following example. Consider the causal graph in Figure 1a where Age and Degree are hidden, and Years refers to years of experience. Suppose now that the company passes a policy that ignores the employee's degree when they decide the salary for entry-level employees, which gives the CSI ($S \perp\!\!\!\perp D \mid Y, J$=*entry-level*). The state-based causal effect $\Pr_{Y\geq 10}(J$=*entry-level*$, S$=*low*$)$ is now identifiable; see Proposition 6 in Appendix D. However, a different state-based causal effect $\Pr_{Y\geq 10}(J$=*senior-level*$, S$=*low*$)$ is unidentifiable since we cannot leverage the CSI when J=*senior-level* (the CSI only takes effect when J=*entry-level*). Therefore, the variable-based causal effect $\Pr_Y(J, S)$ is unidentifiable yet the state-based causal effect $\Pr_{Y\geq 10}(J$=*entry-level*$, S$=*low*$)$ is identifiable.

This example shows that an unidentifiable causal effect at the variable level may be identifiable for a specific treatment and outcome. Hence, existing methods for variable-based identifiability are no longer sufficient for state-based identifiability, highlighting the need for more refined approaches to identify causal effects at this more granular level.

The next question we consider is whether CSI constraints can improve variable-based identifiability like they do for state-based identifiability. That is, if a causal effect is unidentifiable (according to Definition 1), can it become identifiable when CSI constraints are imposed (according to Definition 3)? The answer turns out to be positive. Consider the modified causal graph for the company example depicted in Figure 1b, where Y is no longer a parent of S. In this graph, the variable-based causal effect $\Pr_J(S)$ is unidentifiable. However, suppose we further know that the company determines an employee's job level only based on their age if they have worked for over 10 years — i.e., we have the

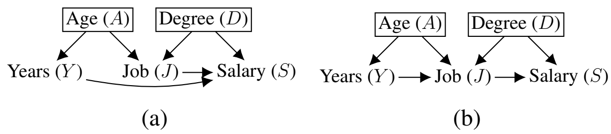


Figure 1: causal graphs showing the effectiveness of CSI constraints, with hidden variables shown in rectangles.

CSI ($J \perp\!\!\!\perp D \mid A, Y \geq 10$). The same causal effect $\Pr_J(S)$ becomes identifiable in this case and can be estimated as $\Pr_J(S) = \Pr(S|J, Y \geq 10)$. We formalize this result as Proposition 7 in Appendix D. The result suggests that classical identification methods (such as IDENTIFY [Tian and Pearl, 2003]) are insufficient for variable-based identifiability when additional CSIs are available.

4 THE IMPACT OF STATE CONSTRAINTS ON IDENTIFIABILITY

We showed earlier that CSI constraints alone can improve both variable-based and state-based identifiability. We next show that these constraints can further improve identifiability when coupled with *state constraints,* which specify the possible states of some variables. Note that CSI constraints do not in principle require the specification of variable states (i.e., set of all states) even for variables that participate as contexts in such constraints. For example, the CSI (Light $\perp\!\!\!\perp$ People | Outage = *yes*) only requires "yes" to be a valid state of the variable Outage but does not require specifying the other states of Outage. Note also that earlier works in [Tikka et al., 2019, Mokhtarian et al., 2022], which studied the impact of CSI constraints on causal effect identifiability, did assume that all variables are binary which is effectively a state constraint. Proposition 8 in Appendix D shows that state constraints have no impact on variable-based or state-based identifiability without additional knowledge (e.g., CSIs), assuming the causal graph is semi-Markovian.[2] This perhaps explains why state constraints did not receive as much attention in the literature on causal effect identifiability. We next consider the interplay between CSIs and state constraints though, showing that the latter can indeed improve both variable-based and state-based identifiability when combined with the former.

Consider the causal graph in Figure 2a which models flight delays. Here, Country (U) refers to the country of departure and is assumed to be hidden. Assume further the CSI constraints ($X \perp\!\!\!\perp U \mid A$=*short*) and ($Y \perp\!\!\!\perp U \mid X, A$=*long*). Without restricting variable states, the variable-based causal effect $\Pr_X(Y)$ is not identifiable. Moreover, no state-based causal effect $\Pr_x(y)$ is identifiable in this case. However, if we assume the states of A to be $\{$*short*, *long*$\}$, the causal effect $\Pr_X(Y)$ becomes identifiable and, hence, also state-based identifiable for any states x, y. In fact, we can estimate the causal effect using the following formula: $\Pr_X(Y) = \Pr(A$=*short*$)$ $\Pr(Y|X, A$=*short*$) + \Pr(A$=*long*$)\Pr(Y|X, A$=*long*$)$. This example is formalized as Proposition 9 in Appendix D. Interestingly, the state constraint of this example involves an observed variable yet it still managed to improve identifiability when coupled with CSIs.

[2]In a Semi-Markovian graph, every hidden variable is a root and has exactly two children [Verma, 1993, Tian and Pearl, 2002]. This assumption is quite common in the existing literature, e.g., the IDENTIFY algorithm in [Tian and Pearl, 2003, Huang and Valtorta, 2006] and the ID algorithm in [Shpitser and Pearl, 2006].

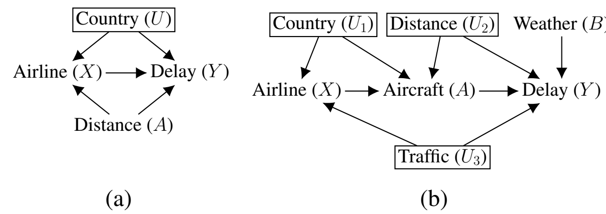


Figure 2: causal graphs showing the effectiveness of state constraints when coupled with CSIs.

The next example shows that adding state constraints to CSIs may enable state-based identifiability but not variable-based identifiability. Consider the causal graph in Figure 2b which depicts a more detailed model of flight delays. Suppose U_1, U_2, U_3 are hidden and we have the following CSIs: ($A \perp\!\!\!\perp U_1 \mid X$=*low-cost*, U_2), ($Y \perp\!\!\!\perp U_3 \mid B$=*raining*, U_2, A), and ($Y \perp\!\!\!\perp A \mid B$=*snowing*,$U_2, U_3$). Without state constraints, the causal effect $\Pr_x(y)$ is unidentifiable for any states x, y (hence, $\Pr_X(Y)$ is not identifiable). If we assume that Weather (B) has the only states $\{$*raining*, *snowing*$\}$, then the causal effect becomes state-based identifiable but still not variable-based identifiable. In particular, the state constraint enables the identifiability of $\Pr_{X=\textit{low-cost}}(Y$=*true*$)$ but not $\Pr_{X=\textit{high-cost}}(Y$=*true*$)$ so $\Pr_X(Y)$ remains unidentifiable. This is formalized as Proposition 10 in Appendix D.

5 TESTING IDENTIFIABILITY UNDER CSI AND STATE CONSTRAINTS

We have so far discussed the role of additional constraints – CSIs and state constraints – in improving both variable-based and state-based identifiability. More importantly, we have demonstrated the separation between these two levels of identifiability, both under CSIs alone and when the two types of constraints are combined. We next introduce a systematic method for testing state-based identifiability in the presence of CSI and state constraints. We start by presenting an algorithm for testing identifiability under CSI constraints only, and then extend it in Section 5.3 to treat state constraints as well. For the case of CSI constraints only, we immediately obtain a method for testing variable-based identifiability by considering specific states of treatments and outcomes based on Proposition 2. For the case with additional state constraints, we can similarly test variable-based identifiability using the method for state-based identifiability as we demonstrate in Section 5.3. One assumption made by our method is that all contexts appearing in the CSIs are observed. This assumption commonly holds when the states of hidden variables are completely unknown (i.e., unconstrained) and therefore cannot appear in the CSI contexts.[3]

Algorithm 1 shows the procedure for testing state-based identifiability, assuming the only constraints are CSIs. The main procedure ID-CSI takes as input the standard components for classical identifiability (a causal graph, observed variables, and treatment and outcome states), along with a set of CSI constraints. The algorithm returns a formula for estimating the causal effect $\Pr_{\mathbf{x}}(\mathbf{y})$ if it is identifiable; otherwise, it returns FAIL. The structure of the algorithm is similar to that of IDENTIFY [Tian and Pearl, 2003]: a causal graph is decomposed into subgraphs, called *c-components*, and the problem of causal effect identification is converted to independent identification problems on these c-components. However, there are two key differences between Algorithm 1 and IDENTIFY. First, Algorithm 1 computes the c-components of a novel notion called *context-induced graph*, which is a pruned version of the causal graph based on CSIs (see Section 5.1). Second, the algorithm utilizes a modified procedure for the identification problem on c-components (of context-induced graphs) which is specified by the ID-QFUNC subroutine (see Section 5.2). A detailed review of c-components and the associated notion of c-factors is provided in Appendix A.

5.1 IDENTIFYING CAUSAL EFFECTS USING CONTEXT-INDUCED GRAPHS

We start with the notion of context-induced graphs (utilized by Algorithm 1), which simplify the causal graph by leveraging CSI constraints. The definition of context-induced graph (CI-graph) takes as input a set of CSIs and a context and removes edges based on the CSIs activated by the context.

Definition 4. *Let G be a causal graph, $\mathcal{C}$ be a set of CSIs, and $\mathbf{s}$ be a context. A CI-graph under $\mathbf{s}$ is obtained from G by repeatedly removing edges as follows: remove edge $P_1 \to C$ if there exists a CSI $(C \perp\!\!\!\perp P_1 \mid \mathbf{p}_2, \mathbf{P}_3)$ for which $\mathbf{p}_2 \subseteq \mathbf{s}$ and $\mathbf{P}_2$ are parents of C at the time of removal.*[4]

Given a fixed context, the choice of CI-graph may not be unique. To illustrate, consider the causal graph in Figure 4a and assume the CSIs $(D \perp\!\!\!\perp B \mid c_0, A)$, $(D \perp\!\!\!\perp C \mid b_0, A)$, and $(E \perp\!\!\!\perp B \mid c_0)$. The graphs in Figures 4b and 4c are both valid CI-graphs under the context b_0c_0. In particular, Figures 4b is obtained by first removing the edge $B \to E$ based on the CSI $(E \perp\!\!\!\perp B \mid c_0)$, and then removing the edge $B \to D$ based on $(D \perp\!\!\!\perp B \mid c_0, A)$. On the other hand, Figures 4b applies the same first step but removes the edge $C \to D$ based on $(D \perp\!\!\!\perp C \mid b_0, A)$ as the second step. No more edges can be removed from either graph. CI-graphs reduce the parameters of a causal graph by removing some of its edges, while preserving the (conditional) independencies implied by the original causal graph and CSI constraints given the context; see Section B for further details.

Instead of testing identifiability using the original causal graph, Algorithm 1 enumerates all possible CI-graphs T under the context $\mathbf{xy}$ and checks whether the causal effect can be identified using any of them (line 2). Similar to the classi-

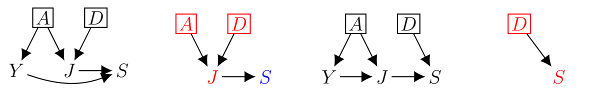
Algorithm 1 State-Based Identifiability with CSI Constraints

1: **procedure** ID-CSI($G, \mathbf{V}, \mathbf{x}, \mathbf{y}, \mathcal{C}$)
// **input:** causal graph G, observed variables $\mathbf{V}$, treatment $\mathbf{x}$, outcome $\mathbf{y}$, CSI constraints $\mathcal{C}$
// **output:** identifying formula for $\Pr_{\mathbf{x}}(\mathbf{y})$ or FAIL
// Assume $\mathbf{V} = (V^1, V^2, \dots)$ are ordered topologically in G, and let $\mathbf{V}^{(i)} = \{V^1, \dots, V^i\}$
2: **for each** CI-graph T under context $\mathbf{xy}$ **do**
3: $M \leftarrow$ remove $\mathbf{X}$ and retain the subgraph of T consisting of all ancestors of $\mathbf{Y}$
4: $\mathbf{C}_1, \dots \mathbf{C}_m \leftarrow$ maximal c-components in M
5: **for** $i = 1$ to m **do**
6: **for each** context $\mathbf{s}$ and its CI-graph H **do**
7: **if** $\mathbf{C}_i$ have the same parents in T and H **and** $(\mathbf{s} \setminus \mathbf{xy})$ contains no state of $\mathbf{C}_i$ or their parents
8: $f^i \leftarrow$ ID-QFUNC($\mathbf{C}_i, \mathbf{V} \cup \mathbf{U}, \Pr(\mathbf{V}), \mathbf{s}, H$) // $\mathbf{U}$ are the hidden variables
9: **if** $f^i \neq$ FAIL **break**
10: **if** $f^i \neq$ FAIL for $i = 1, \dots, m$
11: **return** $\sum_{\mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y})} \prod_{i=1}^m f^i_{\mathbf{xy}}$
12: **return** FAIL

13: **procedure** ID-QFUNC($\mathbf{C}, \mathbf{D}, f, \mathbf{s}, H$)
// **input:** c-component $\mathbf{C}$, variable set $\mathbf{D}$, expression (factor) f, context $\mathbf{s}$, graph H
// **output:** formula for computing the c-factor f^i or FAIL
14: **if** $\mathbf{D} = \mathbf{C}$ **return** $f_{\mathbf{s}}$
15: $H_{\mathbf{D}} \leftarrow$ subgraph of H formed by variables in $\mathbf{D}$
16: **if** exists a leaf node W in $H_{\mathbf{D}}$ and $W \notin \mathbf{S} \cup \mathbf{C}$
17: $\mathbf{D}' \leftarrow \mathbf{D} \setminus \{W\}$
18: $f' \leftarrow \sum_W f$
19: **return** ID-QFUNC($\mathbf{C}, \mathbf{D}', f', \mathbf{s}, H$)
20: **if** exists a maximal c-component $\mathbf{D}'$ in $H_{\mathbf{D}}$ such that $\mathbf{C} \subseteq \mathbf{D}' \subset \mathbf{D}$
21: $f' \leftarrow \prod_{V^i \in (\mathbf{D}' \cap \mathbf{V})} \frac{\sum_{\mathbf{D} \setminus \mathbf{V}^{(i)}} f}{\sum_{\mathbf{D} \setminus \mathbf{V}^{(i-1)}} f}$
22: **return** ID-QFUNC($\mathbf{C}, \mathbf{D}', f', \mathbf{s}, H$)
23: **return** FAIL

[3] This assumption is stronger than [Tikka et al., 2019] which allows hidden contexts, and is weaker than [Mokhtarian et al., 2022] which assumes all contexts are observed roots in the graph.

[4] The notion of CI-graph differs from existing notions that leverage CSIs; see footnote 10 in the appendix for more details.



(a) CI-graph 1a (b) mutilated 1a (c) CI-graph 1b (d) mutilated 1b

Figure 3: context-induced and mutilated graphs for Figure 1.

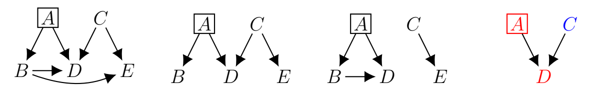
cal identification method IDENTIFY [Tian and Pearl, 2003], the CI-graph T is then mutilated by removing the treatments $\mathbf{X}$ and retaining the subgraph consisting of the ancestors of outcomes $\mathbf{Y}$ (line 3). If we decompose the resulting graph into c-components $\{\mathbf{C}_1, \dots \mathbf{C}_m\}$, we obtain the following key property: $\Pr_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \prod_{i=1}^{m} f^i_{\mathbf{xy}}$, where $f^i_{\mathbf{xy}}$ denotes the entries of the c-factor for $\mathbf{C}_i$ that are compatible with $\mathbf{xy}$.[5] Hence, if all c-factors corresponding to the c-components in the (mutilated) CI-graph are identifiable from the observational distribution $\Pr(\mathbf{V})$, the causal effect is also identifiable. This yields a sound method for testing identifiability by checking identifiability with respect to each independent c-component in the CI-graph. We state this result formally as Lemma 1 with proof in Appendix D.

To illustrate, consider the causal graph in Figure 1a with the CSI $(S \perp\!\!\!\perp D \mid Y, j_0)$ and state-based causal effect $\Pr_{y_0}(j_0, s_0)$. Figures 3a and 3b depict a CI-graph under the context $y_0 j_0 s_0$ (treatment and outcome states) and its mutilated graph, where $\{A, D, J\}$ and $\{S\}$ are the c-components in the mutilated graph (highlighted with colors). Based on our discussion earlier, the causal effect is identifiable if the corresponding c-factors are identifiable. Consider now a different example in Figure 1b with CSI $(S \perp\!\!\!\perp D \mid Y, j_0)$ and causal effect $\Pr_{y_0}(j_0, s_0)$. Figures 3c and 3d depict the CI-graph and its mutilated graph under the context $y_0 j_0 s_0$. Since $\{D, S\}$ is the only c-component in the mutilated graph, the causal effect is identifiable if the c-factor is identifiable. One key observation is that, since CI-graphs are always subgraphs of the original graph, the c-components (and corresponding c-factors) extracted from the mutilated CI-graphs are no larger than those in the original graph, which intuitively allows more causal effects to be identified.

5.2 TREATING C-COMPONENTS OF CI-GRAPHS

We have shown that the problem of identifying causal effects can be converted to that of identifying c-factors induced by the c-components in the CI-graph. We next present a method for testing the identifiability of these c-factors, which involves the ID-QFUNC procedure in Algorithm 1.

[5]The c-factor for each c-component $\mathbf{C}_i$ in graph T is defined as $f^i = \sum_{\mathbf{U}_i} \prod_{C\in\mathbf{C}_i} \theta_{C|\mathbf{P}_C}$, where $\mathbf{U}_i$ denotes the hidden variables in $\mathbf{C}_i$ and $\mathbf{P}_C$ denotes the parents of C in graph T. For example, in Figure 3a, the c-factor for the c-component $\{A, D, J\}$ can be computed as $f = \sum_{AD} \theta_A \theta_D \theta_{J|AD}$, and its entries under the context $y_0 j_0 s_0$ are given by $f_{y_0 j_0 s_0} = \sum_{AD} \theta_A \theta_D \theta_{j_0|AD}$. See Appendix A for additional examples.



(a) original G (b) CI-graph T (c) CI-graph H (d) mutilated T

Figure 4: causal graph, distinct context-induced graphs, and a mutilated graph under the context $b_0 c_0$.

A graph-based method for testing c-factor identifiability was introduced in IDENTIFY [Tian and Pearl, 2003]. This method applies to the c-factors obtained from the original causal graphs but does not extend to c-factors corresponding to the c-components of CI-graphs. We illustrate this limitation with a concrete example in Appendix D Proposition 5. Here, we propose a novel method for treating the identification of c-components (and their corresponding c-factors) within CI-graphs by modifying the approach in IDENTIFY. As we demonstrate later, our method is sound for c-factor identification, and is complete for classical identifiability (without constraints) when combined with the procedure described in Section 5.1.

Our goal is to check the identifiability of the c-factor for each c-component $\mathbf{C}_i$ in the (mutilated) CI-graph T. To achieve this, we start by enumerating all possible contexts $\mathbf{s}$ and their CI-graphs H (line 6). We then check whether (i) variables in $\mathbf{C}_i$ have the same parents in T and H, and (ii) $(\mathbf{s} \setminus \mathbf{xy})$ contains no state of $\mathbf{C}_i$ or of their parents (line 7). The intuition here is that if both conditions are satisfied, the c-factor induced by $\mathbf{C}_i$ in H (denoted $f^i_{\mathbf{s}}$) must contain all the entries of the c-factor of interest (denoted $f^i_{\mathbf{xy}}$).[6] Hence, we can instead check the identifiability of $f^i_{\mathbf{s}}$ by invoking the ID-QFUNC subroutine (line 8). For example, suppose Figures 4b and 4c depict graphs T and H induced by contexts $c_0 e_0$ and $b_0 c_0$. Both conditions hold for the c-component $\{E\}$ since E has the same parent C in the graphs and $(b_0 c_0 \setminus c_0 e_0)$ does not involve E or its parent. In contrast, condition (i) is violated for c-component $\{A, B, D\}$.

The ID-QFUNC checks the identifiability of c-factor $f^i_{\mathbf{s}}$ recursively. It initializes a variable set $\mathbf{D}$ containing all variables $(\mathbf{V} \cup \mathbf{U})$ and iteratively modifies it using the following two operations. Let $H_{\mathbf{D}}$ be the subgraph of H containing variables in $\mathbf{D}$ and all edge between them. At each step, either (1) remove a leaf node in $H_{\mathbf{D}}$ that is not in $\mathbf{S} \cup \mathbf{C}_i$ (lines 16-19); or (2) replace $\mathbf{D}$ by a maximal c-component $\mathbf{D}'$ in $H_{\mathbf{D}}$ such that $\mathbf{D} \supset \mathbf{D}' \supseteq \mathbf{C}_i$ (lines 20-22). If the recursion terminates with $\mathbf{D} = \mathbf{C}_i$, the c-factor is identifiable, and ID-QFUNC returns a formula for its estimation; otherwise, it returns FAIL. We formulate this c-component identification method as Lemma 2 in Appendix D.

[6]The two c-factors contain the same set of CPTs by condition (i). Moreover, if $\mathbf{s}$ contains the state of some variable in the CPTs, that state must appear in $\mathbf{xy}$ by condition (ii), which implies that the entries in $f^i_{\mathbf{s}}$ form a superset of those in $f^i_{\mathbf{xy}}$.

Algorithm 2 State-Based Identifiability with CSI and State Constraints

```
1: procedure ID-CSI-STATE(G, V, x, y, C_1, C_2)
     // input: graph G, observed variables V, treatment x, outcome
     y, CSI constraints C_1, state constraints C_2
     // output: identifying formula for Pr_x(y) or FAIL
2:   for each possible context-partition {s_1, ..., s_k} of V \
     (X ∪ Y) according to C_2 do
3:     for i = 1 to k do
4:       Pr_x(y, s_i) ← ID-CSI(G, C_1, x, y, V)
5:     if Pr_x(y, s_i) ≠ FAIL for i = 1, ..., k
6:       return Σ_{i=1}^{k} Pr_x(y, s_i)
7:   return FAIL
```

We now illustrate how this identifiability method can be applied to treat the c-components in the CI-graph (Section 5.1). Consider again the c-components $\{A, D, J\}$ and $\{S\}$ in the (mutilated) CI-graph in Figure 3b under the context $y_0 j_0 s_0$. To identify these c-components, let j_0 be the context $\mathbf{s}$ in line 6 and Figure 3a be its CI-graph H. Since $\mathbf{s}$ and H satisfy the conditions in line 7, the causal effect is identifiable if ID-QFUNC establishes identifiability for both c-components. This is indeed the case. The first c-component $\{A, D, J\}$ can be obtained by replacing $(\mathbf{V} \cup \mathbf{U})$ with the maximal c-component $\{A, D, J, Y\}$ and then pruning the leaf node Y. The second c-component $\{S\}$ is already a maximal c-component of T.

We now formally state the soundness of Algorithm 1.

Theorem 1. *Let $\mathcal{C}$ be a set of CSI constraints. Then $\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C} \rangle$ if Algorithm 1 returns a formula for estimating $\Pr_{\mathbf{x}}(\mathbf{y})$.*

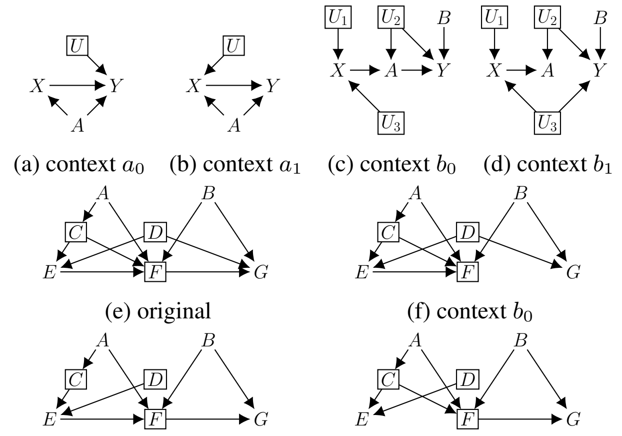
If the set of constraints $\mathcal{C}$ is empty, then Algorithm 1 is both sound and complete for testing identifiability. That is, when $\mathcal{C}$ is empty, Algorithm 1 can identify all causal effects that are identifiable by classical identification methods.

Proposition 3. *$\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable wrt $\langle G, \mathbf{V} \rangle$ iff it is identifiable by Algorithm 1.*

In general, for a causal graph with n variables, the time complexity of Algorithm 1 can be exponential in n, since the numbers of possible CI-graphs (line 2) and contexts (line 6) can be exponential in the worst case. If the total number of contexts appearing in CSIs is bounded by a constant, the worst-case complexity reduces to $O(n^3)$.[7]

We have introduced a sound algorithm for testing state-based (variable-based) identifiability *when the only constraints are CSIs.* When the causal effect is identifiable, the algorithm also returns a formula for estimating the causal effect. We next discuss an extension of this algorithm that further enables leveraging state constraints in conjunction with CSIs.

[7]It suffices to enumerate $\mathbf{s}$ for which all states appear in the contexts of CSIs in line 6 without compromising the completeness. The complexity is dominated by ID-QFUNC in this case. In general, testing causal effect identifiability under CSIs with binary variables is NP-hard [Tikka et al., 2019].



(a) context a_0 (b) context a_1 (c) context b_0 (d) context b_1

(e) original (f) context b_0

(g) context $a_0 b_1$ (h) context $a_1 b_1$

Figure 5: CI-graphs under different contexts in the context-partition for various examples.

5.3 LEVERAGING STATE CONSTRAINTS

As demonstrated in Section 4, additional causal effects become identifiable when CSIs are coupled with state constraints. We next propose a systematic approach for testing state-based (and variable-based) identifiability under both CSI and state constraints by leveraging Algorithm 1. Note that we allow variables with arbitrary cardinalities, which is more general than [Tikka et al., 2019] and [Mokhtarian et al., 2022] which assume all variables are binary.

The detailed procedure is shown in Algorithm 2 which relies on an important notion called *context-partitions*. In particular, we say that a set $\{\mathbf{s}_1, \ldots \mathbf{s}_k\}$ forms a context-partition of variables $\mathbf{V}$ if (i) each $\mathbf{s}_i$, called a context, is an instantiation of some variables in $\mathbf{V}$, (ii) $\mathbf{s}_i$ and $\mathbf{s}_j$ are not compatible for $i \neq j$, and (iii) every instantiation $\mathbf{v}$ of $\mathbf{V}$ is compatible with some context $\mathbf{s}_i$. That is, contexts do not overlap and cover all instantiations of $\mathbf{V}$. To illustrate, suppose variables A, B, C have states $\{a_0, a_1, a_2\}$, $\{b_0, b_1\}$ and $\{c_0, c_1\}$, respectively. Then $\{a_0,\ a_1 b_0,\ a_1 b_1,\ a_2 c_1,\ a_2 c_2\}$ is a context-partition of these variables and so is $\{a_0 b_0,\ a_0 b_1,\ a_1,\ a_2 b_0 c_0,\ a_2 b_0 c_1,\ a_2 b_1\}$.

Algorithm 2 enumerates all possible context-partitions $\{\mathbf{s}_1, \ldots, \mathbf{s}_k\}$ over the variable states specified by state constraints, and reduces the problem of identifying $\Pr_{\mathbf{x}}(\mathbf{y})$ to that of identifying sub-queries $\Pr_{\mathbf{x}}(\mathbf{y}, \mathbf{s}_i)$ for each context $\mathbf{s}_i$ (line 2). It then invokes Algorithm 1 as a subroutine to check the identifiability of these sub-queries (lines 3-6). If all sub-queries are deemed identifiable, $\Pr_{\mathbf{x}}(\mathbf{y})$ is also identifiable. The intuition is that, by augmenting the treatment and outcome states $\mathbf{x}\mathbf{y}$ with additional contexts $\mathbf{s}_i$, we can

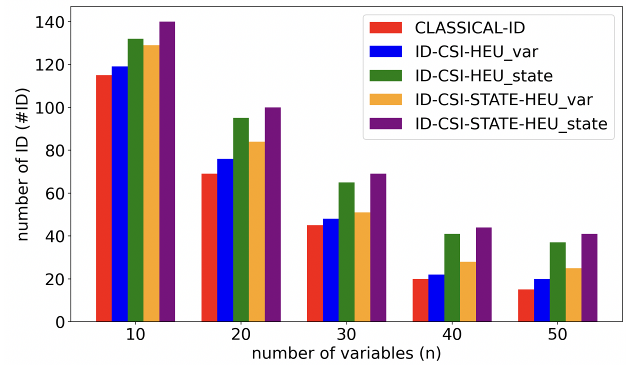


Figure 6: number of identifiable causal effects (out of 200 runs) as the number of variables increases with $r = 0.8$.

exploit additional CSIs whose contexts extend beyond $\mathbf{xy}$ and hence identify more causal effects. When the causal effect is identifiable, the algorithm also returns a formula for estimating it using the equality $\Pr_{\mathbf{x}}(\mathbf{y}) = \sum_i \Pr(\mathbf{y}, \mathbf{s}_i)$.

We now state the soundness of Algorithm 2 for identifiability under CSI and state constraints.

Theorem 2. *Let $\mathcal{C}_1$ be a set of CSIs and $\mathcal{C}_2$ a set of state constraints. Then $\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1 \cup \mathcal{C}_2 \rangle$ if Algorithm 2 returns a formula for estimating $\Pr_{\mathbf{x}}(\mathbf{y})$.*

We next demonstrate how Algorithm 2 can be applied to identify causal effects when state constraints are available. First consider the causal graph in Figure 2a with the CSI constraints $(X \perp\!\!\!\perp U \mid a_0)$, $(Y \perp\!\!\!\perp U \mid X, a_1)$. Suppose variable A only has two states $\{a_0, a_1\}$. Instead of directly checking the identifiability of the original $\Pr_{x_0}(y_0)$, we split the causal effect into sub-queries $\Pr_{x_0}(y_0, a_0)$ and $\Pr_{x_0}(y_0, a_1)$ according to the context-partition $\{a_0, a_1\}$. Figures 5a and 5b depict the CI-graphs for the sub-queries. Since both sub-queries are identifiable by Algorithm 1, the original causal effect is also identifiable; see the proof of Proposition 9 for more details. Similarly, we can apply the theorem to identify $\Pr_{x_0}(y_0)$ in Figure 2b by constructing two sub-queries whose CI-graphs are shown in Figures 5c and 5d; see the proof of Proposition 10 for details.

We now show another example with the causal graph in Figure 5e that requires splitting more than one variable in the context-partition. Suppose the states for variables A, B are $\{a_0, a_1\}, \{b_0, b_1\}$ and assume the following CSI constraints: $(G \perp\!\!\!\perp F \mid b_0, D)$, $(G \perp\!\!\!\perp D \mid b_1, F)$, $(F \perp\!\!\!\perp C \mid a_0, B, E)$, $(F \perp\!\!\!\perp E \mid a_1, B, C)$. The state-based causal effect $\Pr_{e_0}(g_0)$ (and variable-based $\Pr_E(G)$) is identifiable since each sub-query induced by the context-partition $\{b_0, a_0 b_1, a_1 b_1\}$ can be identified using Algorithm 1; see Figures 5f to 5h for the CI-graphs under each context in the partition.

We next discuss a method for testing *variable-based* identifiability under both CSI and state constraints. Similar to the case with CSI constraints only (Proposition 2 in Section 3), the method is based on reducing variable-based identifiability to state-based identifiability. Specifically, to check the identifiability of $\Pr_{\mathbf{X}}(\mathbf{Y})$, we consider instantiations $\mathbf{x}, \mathbf{y}$ in which each state is either included in the state constraints or outside the CSI contexts. The number of such instantiations is finite if every variable outside the state constraints is assigned the same state (outside the CSI contexts) across all instantiations. Then $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable iff $\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable for all these instantiations. We formalize this result as Proposition 11 in Appendix D.

5.4 HEURISTICS

Enumerating CI-graphs (Algorithm 1, line 2), contexts and their corresponding CI-graphs (Algorithm 1, lines 6-7), and context-partitions (Algorithm 2 line 2) incur exponential time complexity and are therefore impractical. To address this issue, we propose in Appendix C heuristics to select a single choice for each step, avoiding the need to enumerate all possible choices. The heuristics reduce the complexity of Algorithm 1 from exponential to polynomial time. Algorithm 2, however, can still require exponential time, since the number of contexts in the context-partition can be exponential in the worst case. Algorithms 1 and 2 combined with these heuristics will be denoted by ID-CSI-HEU and ID-CSI-STATE-HEU. We will use these versions of the algorithms in our upcoming experiments which demonstrate that the heuristics substantially improve the computational efficiency of our proposed identification algorithms.

6 EMPIRICAL RESULTS

We will now present empirical results illustrating the improvement in both variable-based and state-based identifiability due to CSI and state constraints, and further highlighting the separations between the two levels of identifiability.

We consider the following setups of identifiability in our experiments: (1) without constraints (CLASSICAL-ID), (2) with CSI constraints only (ID-CSI-HEU), and (3) with both CSIs and state constraints, assuming all observed variables are binary (ID-CSI-STATE-HEU). For all setups, we evaluate identifiability at both the variable-based level (denoted by the suffix "_var") and state-based level (denoted by the suffix "_state"). Note that the two levels are equivalent in setup (1) based on Proposition 1. For setup (3), we also compare with the DOSEARCH package [Tikka et al., 2019], which targets variable-based identifiability for binary variables only.[8]

For all sets of experiments, we randomly generate 200 causal graphs that contain $n \in \{6, 8, 10, 12, 15, 20, 30, 40, 50\}$ variables. Each variable has at most 5 parents. Among the variables, 70% are designated to be observed, 20% as treat-

[8]The DOSEARCH package is available at https://github.com/santikka/dosearch.

ment, and 20% as outcome. The CSI constraints are randomly generated as follows. For each variable C with parent P_1, we assign a CSI $(C \perp\!\!\!\perp P_1 \mid \mathbf{p}_2, \mathbf{P}_3)$ with probability $r \in \{0.2, 0.5, 0.8\}$. We then assign a random state of other observed parents to $\mathbf{p}_2$ with probability 0.5. All remaining parents are assigned to $\mathbf{P}_3$. Note that larger values of r correspond to more imposed CSIs.

Table 1 in the Appendix reports the average number of identifiable causal effects and runtime for each method, with a 5-minute timeout for each run. Figure 6 plots the number of identifiable causal effects with an increasing number of variables n and a fixed r=0.8. The following patterns can be observed. First, for both variable-based and state-based identifiability, more causal effects become identifiable when additional CSI constraints are imposed (i.e., larger r) and when state constraints are present, which matches our intuition that constraints can facilitate identifiability. Second, the gap between state-based and variable-based identifiability widens as more CSI constraints are imposed. For example, when n=50 and state constraints are present, the gap in the number of identifiable causal effects increases from 4 to 16 when r increases from 0.2 to 0.8. A similar trend holds in the absence of state constraints. Finally, our algorithm ID-CSI-STATE-HEU identifies a comparable – sometimes larger – number of variable-based causal effects compared to DOSEARCH, while achieving significantly higher computational efficiency (DOSEARCH times out on the majority of runs when $n \geq 15$). These results demonstrate the effectiveness of our proposed algorithms when applied to both state-based and variable-based identifiability.

It is worth emphasizing that our algorithms adopt a graph decomposition approach using c-components, in contrast to the exhaustive search method based on CSI-calculus rules introduced in [Tikka et al., 2019] and used in DOSEARCH. In fact, one may draw an interesting analogy here: our algorithms relate to CSI-calculus in a similar way to how IDENTIFY [Tian and Pearl, 2003] relate to the do-calculus [Pearl, 2009] as noted in [Tikka et al., 2019]. A key advantage of our procedural approach is that it makes the identification process more explicit than exhaustive search, enabling computational optimizations (e.g., the heuristics developed in Section 5.4) and facilitating future analysis of completeness.

7 CONCLUSION

We considered how the identifiability of causal effects can be improved by focusing on a particular treatment and outcome and by including additional knowledge beyond causal graphs. This led us to suggest the more granular notion of state-based identifiability. It also led to some findings on the extent to which various types of knowledge and their combinations can draw distinctions between state-based and variable-based identifiability. Another contribution of this work is the introduction of novel algorithms for both levels of causal effect identification under additional CSI and state constraints, which we evaluated empirically.

References

Craig Boutilier, Nir Friedman, Moisés Goldszmidt, and Daphne Koller. Context-specific independence in bayesian networks. In *UAI*, pages 115–123. Morgan Kaufmann, 1996.

Yizuo Chen and Adnan Darwiche. Identifying causal effects under functional dependencies. *Entropy*, 26(12), 2024. ISSN 1099-4300.

Yizuo Chen and Adnan Darwiche. Constrained identifiability of causal effects. *AAAI Workshop on Artificial Intelligence with Causal Techniques*, 2025.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If.* Boca Raton: Chapman & Hall/CRC, 2020.

Yimin Huang and Marco Valtorta. Identifiability in causal bayesian networks: A sound and complete algorithm. In *AAAI*, pages 1149–1154. AAAI Press, 2006.

Jaron Jia Rong Lee, AmirEmad Ghassami, and Ilya Shpitser. A general identification algorithm for data fusion problems under systematic selection. In *UAI*, volume 244 of *Proceedings of Machine Learning Research*, pages 2188–2204. PMLR, 2024.

Ehsan Mokhtarian, Fateme Jamshidi, Jalal Etesami, and Negar Kiyavash. Causal effect identification with context-specific independence relations of control variables. In *AISTATS*, volume 151 of *Proceedings of Machine Learning Research*, pages 11237–11246. PMLR, 2022.

Judea Pearl. [bayesian analysis in expert systems]: Comment: Graphical models, causality and intervention. *Statistical Science*, 8(3):266–269, 1993.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995. ISSN 00063444.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, second edition, 2009.

Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

Johan Pensar, Henrik J. Nyman, Timo Koski, and Jukka Corander. Labeled directed acyclic graphs: a generalization of context-specific independence in directed graphical models. *Data Min. Knowl. Discov.*, 29(2):503–533, 2015.

Yujia Shen, Arthur Choi, and Adnan Darwiche. A new perspective on learning context-specific independence. In *PGM*, volume 138 of *Proceedings of Machine Learning Research*, pages 425–436. PMLR, 2020.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *AAAI*, pages 1219–1226. AAAI Press, 2006.

Ilya Shpitser and Judea Pearl. Complete identification methods for the causal hierarchy. *J. Mach. Learn. Res.*, 9: 1941–1979, 2008.

Jin Tian and Judea Pearl. On the testable implications of causal models with hidden variables. In *UAI*, pages 519–527. Morgan Kaufmann, 2002.

Jin Tian and Judea Pearl. On the identification of causal effects. *Technical Report*, R-290-L, 2003.

Santtu Tikka, Antti Hyttinen, and Juha Karvanen. Identifying causal effects via context-specific independence relations. In *NeurIPS*, pages 2800–2810, 2019.

Thomas Verma and Judea Pearl. Causal networks: semantics and expressiveness. In *UAI*, pages 69–78. North-Holland, 1988.

Thomas S. Verma. Graphical aspects of causal models. *Technical Report*, R-191, 1993.

A BACKGROUND ON C-COMPONENTS AND C-FACTORS

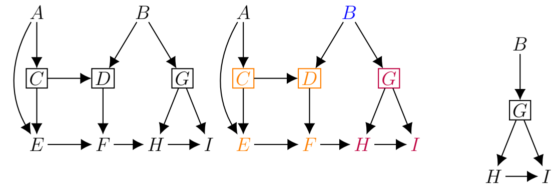
Sound and complete methods for testing classical identifiability based on causal graphs have been developed in the past; see, e.g., IDENTIFY [Tian and Pearl, 2003, Huang and Valtorta, 2006] and ID [Shpitser and Pearl, 2006]. These methods all rely on a key graphical notion called *c-component*, which we review next.

Definition 5 (c-component [Tian and Pearl, 2002])**.** *Let G be a causal graph with a subset of observed variables. A variable set $\mathbf{C}$ is said to be a c-component in G if it satisfies the following conditions: (i) every pair of hidden variables in $\mathbf{C}$ are either adjacent or have a common child; and (ii) every observed variable in $\mathbf{C}$ has some hidden parent in $\mathbf{C}$.*

A c-component is *maximal* if it is not a subset of any other c-component. We can always *partition* the variables (both observed and hidden) in a causal graph into maximal c-components. For example, Figure 7b depicts the c-component partition of the causal graph, where variables in different c-components are highlighted with different colors.

Moreover, each c-component $\mathbf{C}$ induces a *c-factor* defined as $Q[\mathbf{C}] = \sum_{\mathbf{H}} \prod_{C \in \mathbf{C}} \theta_{C|\mathbf{P}_C}$, where $\mathbf{H}$ denotes the hidden variables in $\mathbf{C}$ and $\mathbf{P}_C$ denotes the parents of C in G. Based on the definition, the c-factor $Q[\mathbf{C}]$ is over all observed variables appearing in the CPTs of $\mathbf{C}$. To illustrate, the c-factor for the c-component $\{C, D, E, F\}$ in Figure 7b is computed as $Q[CDEF] = \sum_{CD} \theta_{C|A}\theta_{D|BC}\theta_{E|AC}\theta_{F|DE}$, which is over variables $\{A, B, E, F\}$. One important property of c-factors is that they decompose the observational distribution $\Pr(\mathbf{V})$ as follows. Let $\{\mathbf{C}_1, \ldots, \mathbf{C}_m\}$ be the c-components that partition variables in G, then $\Pr(\mathbf{V}) = \prod_{i=1}^{m} Q[\mathbf{C}_i]$ [Tian and Pearl, 2003]. In Figure 7b, for example, $\Pr(\mathbf{V}) = Q[A]Q[B]Q[CDEF]Q[GHI]$. In this work, we will apply this c-component (c-factor) decomposition technique for causal effect identification under constraints.

It is standard to convert the problem of identifying causal effects to independent problems of identifying c-factors as shown in the IDENTIFY algorithm [Tian and Pearl, 2003]. In particular, a c-factor is identifiable iff its entries are uniquely determined (computable) by the observational distribution $\Pr(\mathbf{V})$. Consider a causal graph G, treatments $\mathbf{X}$, and outcomes $\mathbf{Y}$. Let T be the mutilated graph obtained by first removing treatment variables $\mathbf{X}$ from G and then retaining only the ancestors of $\mathbf{Y}$ (Figure 7c depicts the mutilated graph of Figure 7a). The IDENTIFY algorithm then reduces the identifiability of the causal effect $\Pr_{\mathbf{X}}(\mathbf{Y})$ to that of c-factors induced by the c-components in T. This reduction is both sound and complete [Huang and Valtorta, 2006, Shpitser and Pearl, 2006]: the causal effect is variable-based (state-based) identifiable iff all the c-factors are identifiable.



(a) original G (b) c-components (c) mutilated

Figure 7: The causal graph G, its mutilated graph under $\Pr_F(I)$, and the c-component decomposition. Hidden variables are shown in rectangles, and variables in different c-components are highlighted with different colors.

B PROPERTIES OF CONTEXT-INDUCED GRAPHS

Context-induced graphs (CI-graphs) introduced in Definition 4 satisfy two key properties which we formalize in the following proposition.

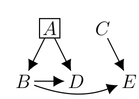
Proposition 4. *Consider a causal graph G and a set of CSIs $\mathcal{C}$. Let T be a CI-graph under context $\mathbf{s}$. Then the following properties hold for all models Θ of G satisfying $\mathcal{C}$:*

(i) *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{S}$ be disjoint sets of variables. If $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $(\mathbf{Z} \cup \mathbf{S})$ in T, then $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}, \mathbf{s})$ in the distribution induced by Θ.*[9]

(ii) *For each variable V with parents $\mathbf{P}$ in G and parents $\mathbf{P}'$ in T, $\theta_{v|\mathbf{p}_1} = \theta_{v|\mathbf{p}_2}$ for all instantiations $\mathbf{p}_1, \mathbf{p}_2$ of $\mathbf{P}$ that extend a common $\mathbf{p}'$ that is compatible with $\mathbf{s}$.*[10]

The first property shows that CI-graphs preserve the conditional independencies given the context $\mathbf{s}$, and the second property implies that the parameters in CI-graphs characterize the equal parameters under the CSIs. In particular, we can always construct a model for the CI-graph by setting each parameter $\theta_{v|\mathbf{p}'}$ equal to $\theta_{v|\mathbf{p}}$ for any $\mathbf{p} \supseteq \mathbf{p}'$ (the value of $\theta_{v|\mathbf{p}'}$ is unique if $\mathbf{p}'$ is compatible with $\mathbf{s}$). The resulting model satisfies several important properties, including the preservation of $\Pr(\mathbf{v})$ for all $\mathbf{v}$ compatible with the context $\mathbf{s}$. We will use this construction technique to prove the soundness of our identification method (Algorithm 1).

We next illustrate the properties in Proposition 4 using the following example. The figure on the right depicts a CI-graph of Figure 4a under context b_0, assuming CSIs



[9] See [Verma and Pearl, 1988] for details of d-separation and its interplay with conditional independencies.

[10] Let $\mathbf{x}, \mathbf{y}$ be two instantiations. We say that $\mathbf{x}$ *extends* $\mathbf{y}$ if $\mathbf{x}$ contains all the states in $\mathbf{y}$, and that $\mathbf{x}$ and $\mathbf{y}$ are *compatible* if they do not contain conflicting states. The CI-graphs considered here differ from the notion of *context-specific graph* in [Pensar et al., 2015, Tikka et al., 2019, Lee et al., 2024]: the latter satisfies property (i) but not property (ii). Moreover, unlike context-specific graphs, CI-graphs are not necessarily unique.

$(D \perp\!\!\!\perp B \mid c_0, A)$, $(D \perp\!\!\!\perp C \mid b_0, A)$, and $(E \perp\!\!\!\perp B \mid c_0)$. Since A and E are d-separated by B in the CI-graph, $(A \perp\!\!\!\perp E \mid b_0)$ by property (i). Moreover, $\theta_{D|A,b_0,c_1} = \theta_{D|A,b_0,c_2}$ holds for any states c_1, c_2 of C by property (ii).

C HEURISTICS

We present heuristics that select a single choice for the CI-graph under $\mathbf{xy}$ (Algorithm 1, line 2), the context $\mathbf{s}$ and its CI-graph (Algorithm 1, lines 6-7), and the context-partition (Algorithm 2 line 2) when state constraints are imposed. Although the use of heuristics may compromise the completeness of the algorithms, it makes them computationally feasible by avoiding exponential enumerations and still substantially improves the identifiability of causal effects, as demonstrated by the empirical results in Section 6.

CI-graph under $\mathbf{xy}$ (Algorithm 1, line 2). We begin with a reverse-topological ordering Π over states $\mathbf{xy}$. Let $\mathbf{s}$ be a context initialized to be empty, and T be a graph initialized as the original graph G. We iterate over each state in Π in order, adding it to $\mathbf{s}$ and removing from T all edges enabled by the CSIs under the updated $\mathbf{s}$. The resulting graph T is guaranteed to be a CI-graph for $\mathbf{xy}$, and no additional edges can be pruned from the CI-graph after the procedure.

Context $\mathbf{s}$ and its CI-graph H (Algorithm 1, lines 6-7). The intuition is to activate CSIs and remove edges one at a time, returning immediately if the CI-graph can be used to establish identifiability of the target c-component $\mathbf{C}_i$.

We initialize the context $\mathbf{s}$ as follows. For each variable W that is in $\mathbf{C}_i$ or is a parent of some variable in $\mathbf{C}_i$ and has a state $w \in \mathbf{xy}$, we add w to $\mathbf{s}$. We initialize H as the original graph G. For each variable $C \in \mathbf{C}_i$ with some parent P, we remove the edge $P \rightarrow C$ from H if the edge is not in T. The initialized $\mathbf{s}$ and H satisfy the condition in line 7.

We now augment the context $\mathbf{s}$ and prune additional edges from H in an iterative way. At each iteration, let $\mathbf{C} \supseteq \mathbf{C}_i$ be the c-component of H resulting from ID-QFUNC. If $\mathbf{C} = \mathbf{C}_i$, we terminate the iterative process since the c-factor of $\mathbf{C}_i$ is identifiable under $\mathbf{s}$ and H. Otherwise, we pick a CSI $(W \perp\!\!\!\perp P_1 \mid \mathbf{p}_2, \mathbf{P}_3)$ that satisfies the following conditions: $W \in (\mathbf{C} \setminus \mathbf{C}_i)$, $P_1 \in \mathbf{C}$, and $\mathbf{p}_2$ is compatible with $\mathbf{s}$. When multiple such CSIs exist, we prioritize the one whose P_1 is hidden and has fewest children in $H_{\mathbf{C}}$, where $H_{\mathbf{C}}$ denotes the subgraph of H only consisting of $\mathbf{C}$. We then add states $\mathbf{p}_2$ to the context $\mathbf{s}$ and remove the edge $P_1 \rightarrow C$ from H, before proceeding to the next iteration. The process terminates if no CSI satisfies the above conditions.

Context-partition (Algorithm 2 line 2). The intuition is to include as few variables as possible in the context-partitions, as long as doing so is unlikely to compromise identifiability. In particular, we construct the contexts in the partition and their corresponding CI-graphs simultaneously. We begin with a reverse-topological ordering Π over observed variables $\mathbf{V}$. Let $\mathbf{s}$ be a context initialized to be empty, and T be a graph initialized as the original graph G.

We iterate over each variable V in Π in order. At each iteration, let M be the mutilated graph obtained by removing $\mathbf{X}$ and retaining the ancestors of $(\mathbf{Y} \cup \mathbf{S})$ in T. If $V \in \mathbf{X} \cup \mathbf{Y}$ with state $v \in \mathbf{xy}$, we remove from T all edges enabled by the CSIs under the context $\mathbf{s} \cup \{v\}$, and continue the iteration using the current context $\mathbf{s}$ and the pruned graph.

If $V \notin \mathbf{X} \cup \mathbf{Y}$ but every state of V appears in the context of some CSI in M, we create a new context $\mathbf{s} \cup \{v_i\}$ for each state v_i of $\mathbf{V}$. For each new context, we assign a graph obtained by removing from T all edges enable by the CSIs under that new context. We then continue the iteration for each new context and its assigned graph.

Otherwise, we skip V and proceed with the current context $\mathbf{s}$ and graph T. After all variables have been processed, we collect the resulting contexts to form the context-partition and take their associated graphs as the CI-graphs.

D STATEMENTS AND PROOFS

The statements and proofs are presented in a different order from that of the main paper. We begin with the soundness proofs of the algorithms introduced in Section 5 and then establish the separation results in Sections 3 and 4.

D.1 SOUNDNESS PROOFS

Proof of Proposition 4. We first show that property (ii) holds after each edge removal in the definition of CI-graph (Definition 4). Initially, $\mathbf{P}' = \mathbf{P}$ and hence $\theta_{v|\mathbf{p}_1} = \theta_{v|\mathbf{p}_2}$ for $\mathbf{p}_1 = \mathbf{p}_2$. Suppose now the property holds for graph T before the removal of edge $P_1 \rightarrow C$ based on the CSI $(C \perp\!\!\!\perp P_1 \mid \mathbf{p}_2, \mathbf{P}_3)$. We want to show that property (ii) holds after the edge removal. Let $\mathbf{P}$ be the parents of C in T and $\mathbf{P}'$ be the parents after removal. Let $\mathbf{p}^1, \mathbf{p}^2$ be any two instantiations extending a common $\mathbf{p}'$ compatible with $\mathbf{s}$.

Suppose C appears in the context, i.e., $c^\star \in \mathbf{s}$ for some state $c^\star$. Then $\theta_{v|\mathbf{p}^1} = \theta_{v|\mathbf{p}^1_{c^\star}}$ and $\theta_{v|\mathbf{p}^2} = \theta_{v|\mathbf{p}^2_{c^\star}}$ according to the CSI constraint, where $\mathbf{p}^1_{c^\star}$, $\mathbf{p}^2_{c^\star}$ are obtained from $\mathbf{p}_1, \mathbf{p}_2$ by replacing the state of C by $c^\star$. It follows that $\mathbf{p}^1_{c^\star}$, $\mathbf{p}^2_{c^\star}$ extend a common $\mathbf{p}$ that is compatible with $\mathbf{s}$. Hence, $\theta_{v|\mathbf{p}^1} = \theta_{v|\mathbf{p}^1_{c^\star}} = \theta_{v|\mathbf{p}^2_{c^\star}} = \theta_{v|\mathbf{p}^2}$ by induction.

Suppose C does not appear in the context, i.e., $C \notin \mathbf{S}$. Let $c^\star$ be an arbitrary state of C. Then $\theta_{v|\mathbf{p}^1} = \theta_{v|\mathbf{p}^1_{c^\star}}$ and $\theta_{v|\mathbf{p}^2} = \theta_{v|\mathbf{p}^2_{c^\star}}$ according to the CSI constraint. Since both $\mathbf{p}^1_{c^\star}$ and $\mathbf{p}^2_{c^\star}$ are compatible with $\mathbf{s}$, $\theta_{v|\mathbf{p}^1} = \theta_{v|\mathbf{p}^1_{c^\star}} = \theta_{v|\mathbf{p}^2_{c^\star}} = \theta_{v|\mathbf{p}^2}$ by induction.

We next prove property (i). For any model Θ for G that

(a) Original G (b) CI-graph T

Figure 8: CI-graph constructed for the causal effect $\Pr_x(y)$ and CSI constraint $(Z \perp\!\!\!\perp U \mid x)$.

induces $\Pr$, we want to construct a model for T that induces a distribution $\Pr'$ satisfying $\Pr(\mathbf{v}) = \Pr'(\mathbf{v})$ for all instantiations $\mathbf{v} \supseteq \mathbf{s}$. This ensures that if d-separation holds, i.e., $\Pr'(\mathbf{X}|\mathbf{Y}, \mathbf{Z}, \mathbf{s}) = \Pr'(\mathbf{X}|\mathbf{Z}, \mathbf{s})$, the conditional independency also holds, i.e., $\Pr(\mathbf{X}|\mathbf{Y}, \mathbf{Z}, \mathbf{s}) = \Pr(\mathbf{X}|\mathbf{Z}, \mathbf{s})$. The construction can be done as follows. For each variable C with parents $\mathbf{P}'$ in T, we define its CPT entries as $\theta'_{c|\mathbf{p}'} = \theta_{c|\mathbf{p}}$ for every instantiation $(c, \mathbf{p}')$, where $\mathbf{p}$ can be any instantiation where $\mathbf{p} \supseteq \mathbf{p}'$. By property (ii), when $\mathbf{p}'$ is compatible with $\mathbf{s}$, the value of $\theta'_{c|\mathbf{p}'}$ is independent of the choice of $\mathbf{p}$ since all the $\theta_{c|\mathbf{p}}$ are equal. Hence, $\Pr(\mathbf{v}) = \Pr'(\mathbf{v})$ for all $\mathbf{v}$ compatible with $\mathbf{s}$. □

Before showing the soundness proofs for the algorithms, we first clarify the models (parameterizations) for CI-graphs so that the distributions induced by CI-graphs are well-defined. Let T be a CI-graph of the original graph G under some context $\mathbf{s}$ by Definition 4. We assume that the model Θ' of T is constructed from the model Θ of G as follows. For each variable C with parents $\mathbf{P}$ in G and parents $\mathbf{P}'$ in T, we assign $\theta'_{c|\mathbf{p}'} = \theta_{c|\mathbf{p}}$ for every instantiation $(c, \mathbf{p}')$, where $\mathbf{p}$ can be any instantiation satisfying $\mathbf{p} \supseteq \mathbf{p}'$; this is the same as the construction method shown in Appendix B. Based on the construction, we can now establish some key results on the c-factors induced by the CI-graph.

Lemma 1. *Consider a CI-graph T under the context $\mathbf{xy}$. Let M be the mutilated graph constructed in Algorithm 1 line 3 and let $\{\mathbf{C}_1, \ldots, \mathbf{C}_m\}$ denote its maximal c-components. The causal effect $\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable if the c-factors $Q[\mathbf{C}_i]_{\mathbf{xy}}$ in T are identifiable for all $i = 1, \ldots, m$.*

Proof. Let Θ be the model for the original graph G and Θ' be the model for the CI-graph T constructed using the above procedure. Let $\Pr$ and $\Pr'$ be the distributions induced by Θ and Θ', respectively. By Proposition 4, Θ' agrees with Θ on all CPT entries that are compatible with $\mathbf{xy}$. As a result, $\Pr'_{\mathbf{x}}(\mathbf{v}) = \Pr_{\mathbf{x}}(\mathbf{v})$ for all instantiations $\mathbf{v} \supseteq \mathbf{y}$. Therefore, the CI-graph T can be used for causal identification since it preserves the value of causal effect, i.e., $\Pr'_{\mathbf{x}}(\mathbf{y}) = \Pr_{\mathbf{x}}(\mathbf{y})$. The techniques of graph mutilation and c-component (c-factor) decomposition follow from classical identification methods such as IDENTIFY [Tian and Pearl, 2003]. □

Proposition 5. *There exists a graph G and its CI-graph T under context $\mathbf{xy}$ for which $\Pr_{\mathbf{x}}(\mathbf{y})$ is unidentifiable wrt $\langle G, \mathbf{V}, \mathcal{C}\rangle$ but identifiable wrt $\langle T, \mathbf{V}\rangle$.*

Proof. Consider the causal graph G in Figure 8a with the CSI constraint $(Z \perp\!\!\!\perp U \mid x)$. To compute the causal effect $\Pr_x(y)$, we construct the CI-graph T under xy as shown in Figure 8b. Applying classical identification methods (e.g., the front-door criterion Pearl [1995]) to T would suggest that the causal effect is identifiable. However, this conclusion is incorrect: we can construct the following parameterizations Θ^1, Θ^2 for G that satisfy the CSI constraint and induce the same $\Pr(\mathbf{V})$ yet yield distinct $\Pr_x(y)$.

U	θ^1_U	θ^2_U
u	0.5	0.4
$\bar{u}$	0.5	0.6

U	X	$\theta^1_{X\|U}$	$\theta^2_{X\|U}$
u	x	0.6	0.25
u	$\bar{x}$	0.4	0.75
$\bar{u}$	x	0.2	0.5
$\bar{u}$	$\bar{x}$	0.8	0.5

U	X	Z	$\theta^1_{Z\|U,X}$	$\theta^2_{Z\|U,X}$
u	x	z	0.5	0.5
u	x	$\bar{z}$	0.5	0.5
u	$\bar{x}$	z	0.3	1/3
u	$\bar{x}$	$\bar{z}$	0.7	2/3
$\bar{u}$	x	z	0.5	0.5
$\bar{u}$	x	$\bar{z}$	0.5	0.5
$\bar{u}$	$\bar{x}$	z	0.6	2/3
$\bar{u}$	$\bar{x}$	$\bar{z}$	0.4	1/3

U	Z	Y	$\theta^1_{Y\|U,Z}$	$\theta^2_{Y\|U,Z}$
u	z	y	0.5	0.76
u	z	$\bar{y}$	0.5	0.24
u	$\bar{z}$	y	0.8	0.42
u	$\bar{z}$	$\bar{y}$	0.2	0.58
$\bar{u}$	z	y	0.55	0.43
$\bar{u}$	z	$\bar{y}$	0.45	0.57
$\bar{u}$	$\bar{z}$	y	0.3	0.76
$\bar{u}$	$\bar{z}$	$\bar{y}$	0.7	0.24

□

Lemma 2. *Let T be a CI-graph of the original graph G under some context $\mathbf{s}$. For each c-component $\mathbf{C}$ in T (not necessarily maximal), its corresponding c-factor $Q[\mathbf{C}]_{\mathbf{s}}$ is identifiable if $\mathbf{C}$ can be obtained from T by repeatedly applying the following operations:*[11]

(i). Remove a leaf node that is not in $\mathbf{S}$; or

(ii). Replace T with the subgraph formed by one of its maximal c-components.

Proof. Consider again the construction method above Lemma 1. Let Θ and Θ' be the parameterizations for the original graph G and CI-graph T. Let $\Pr$ and $\Pr'$ be the distributions induced by Θ and Θ'. By Proposition 4 property (ii), the equality $\Pr(\mathbf{u}, \mathbf{v}) = \Pr'(\mathbf{u}, \mathbf{v})$ holds for all instantiations $(\mathbf{v}, \mathbf{u})$ compatible with $\mathbf{s}$.

[11] For simplicity, we use T to refer to the graph and its set of variables interchangeably.

We first extend the definition of c-factors to any variable sets $\mathbf{W}$ (not necessarily c-components): $Q[\mathbf{W}] = \sum_{\mathbf{U_W}} \prod_{W \in \mathbf{W}} \theta_{W|\mathbf{P}_W}$, where $\mathbf{U_W}$ are the hidden variables in $\mathbf{W}$ and $\mathbf{P}_W$ contains the parents of W in T. When $\mathbf{W}$ is a c-component in T, the definition is identical to the one in Appendix A. Moreover, we use $Q[\mathbf{W}]_\mathbf{s}$ to denote the instantiations of $Q[\mathbf{W}]$ compatible with the context $\mathbf{s}$.

Our goal is to show that the c-factors obtained from the two operations in this lemma are identifiable (computable from $\Pr(\mathbf{V})$). Let $\mathbf{W}$ be the result of applying the two operations. We prove the following statements by induction: (i) $Q[\mathbf{W}]_\mathbf{s}$ is identifiable; and (ii) $Q[\mathbf{W}^{(i)}]_\mathbf{s}$ is identifiable for each observed $W^i \in \mathbf{W}$, where W^i is the i^{th} variable in $\mathbf{W}$ based on some fixed topological order Π of G, $\mathbf{W}^{(i)}$ denotes all variables ordered before and including W^i, and $Q[\mathbf{W}^{(i)}]$ is defined as $\sum_{\mathbf{W} \setminus \mathbf{W}^{(i)}} Q[\mathbf{W}]$ Tian and Pearl [2002].

We start with the base case where $\mathbf{W}$ contains all variables in T, i.e., $\mathbf{W} = \mathbf{V} \cup \mathbf{U}$. By construction, $Q[\mathbf{W}]_\mathbf{s} = \Pr'(\mathbf{V})_\mathbf{s} = \Pr(\mathbf{V})_\mathbf{s}$ and hence statement (i) holds. Now consider statement (ii). For each observed $W^i \in \mathbf{W}$, all variables in $(\mathbf{W} \setminus \mathbf{W}^{(i)})$ can be pruned from both G and T. The CPT entries for the remaining variables $\mathbf{W}^{(i)}$ agree on all instantiations compatible with $\mathbf{s}$. Hence, $Q[\mathbf{W}^{(i)}]_\mathbf{s} = \Pr'(\mathbf{W}^{(i)} \cap \mathbf{V})_\mathbf{s} = \Pr(\mathbf{W}^{(i)} \cap \mathbf{V})_\mathbf{s}$ which is identifiable.

Suppose both statements hold for some variable set $\mathbf{W}$. Suppose a leaf $L \notin \mathbf{S}$ is pruned from $\mathbf{W}$. Let $\mathbf{D}$ be the remaining variables, then the resulting c-factor $Q[\mathbf{D}]_\mathbf{s}$ is

$$Q[\mathbf{D}]_\mathbf{s} = (\sum_L Q[\mathbf{W}])_\mathbf{s} = \sum_L Q[\mathbf{W}]_\mathbf{s}$$

By induction, the identifiability of $Q[\mathbf{D}]_\mathbf{s}$ follows from the identifiability $Q[\mathbf{W}]_\mathbf{s}$.

We next show that $Q[\mathbf{D}^{(i)}]$ is identifiable for each observed $D^i \in \mathbf{D}$ ordered based on Π. WLG, suppose D^i is ranked as the j^{th} variable in $\mathbf{W}$, i.e., $D^i = W^j$. We consider two cases based on the location of the pruned leaf L. If $L \in \mathbf{W}^{(j)}$, then

$$Q[\mathbf{D}^{(i)}]_\mathbf{s} = (\sum_L \sum_{\mathbf{W} \setminus \mathbf{W}^{(j)}} Q[\mathbf{W}])_\mathbf{s} = \sum_L Q[\mathbf{W}^{(j)}]_\mathbf{s}$$

which can be computed by summing out L from $Q[\mathbf{W}^{(j)}]_\mathbf{s}$.

If $L \notin \mathbf{W}^{(j)}$, then

$$Q[\mathbf{D}^{(i)}]_\mathbf{s} = (\sum_{\mathbf{W} \setminus \mathbf{W}^{(j)}} Q[\mathbf{W}])_\mathbf{s} = Q[\mathbf{W}^{(j)}]_\mathbf{s}$$

which is also equivalent to summing out L from $Q[\mathbf{W}^{(j)}]_\mathbf{s}$. Since $Q[\mathbf{W}^{(j)}]_\mathbf{s}$ is identifiable by induction, $Q[\mathbf{D}^{(i)}]_\mathbf{s}$ is also identifiable in both cases.

Suppose now $\mathbf{W}$ is replaced by a maximal c-component $\mathbf{D} \subset \mathbf{W}$. According to [Tian and Pearl, 2003, Lemma 4], the c-factor for $\mathbf{D}$ can be computed as $Q[\mathbf{D}] = \prod_{W^i \in \mathbf{D} \cap \mathbf{V}} \frac{Q[\mathbf{W}^{(i)}]}{Q[\mathbf{W}^{(i-1)}]}$. Since each $Q[\mathbf{W}^{(i)}]_\mathbf{s}$ is identifiable by induction, $Q[\mathbf{D}]_\mathbf{s}$ is also identifiable. We next show that $Q[\mathbf{D}^{(i)}]_\mathbf{s}$ is identifiable for each observed $D^i \in \mathbf{D}$. By definition, $Q[\mathbf{D}^{(i)}] = \sum_{\mathbf{D} \setminus \mathbf{D}^{(i)}} Q[\mathbf{D}]$. To simplify the expression further, we iteratively eliminate observed variables in $(\mathbf{D} \setminus \mathbf{D}^{(i)})$ in reverse of Π (we do not need to eliminate hidden variables since they have already been summed out in $Q[\mathbf{D}]$). WLG, let W^j be the first observed variable in the reverse order, then

$$\begin{aligned}
\sum_{W^j} Q[\mathbf{D}] &= \sum_{W^j} \prod_{W^k \in \mathbf{D} \cap \mathbf{V}} \frac{Q[\mathbf{W}^{(k)}]}{Q[\mathbf{W}^{(k-1)}]} \\
&= (\prod_{W^k \in \mathbf{D} \cap \mathbf{V}, k<j} \frac{Q[\mathbf{W}^{(k)}]}{Q[\mathbf{W}^{(k-1)}]})(\sum_{W^j} \frac{Q[\mathbf{W}^{(j)}]}{Q[\mathbf{W}^{(j-1)}]}) \\
&= \prod_{W^k \in \mathbf{D} \cap \mathbf{V}, k<j} \frac{Q[\mathbf{W}^{(k)}]}{Q[\mathbf{W}^{(k-1)}]}
\end{aligned}$$

The second equality holds since W^j appears only in $\mathbf{W}^{(j)}$. We can apply the same procedure to eliminate all variables in $(\mathbf{D} \setminus \mathbf{D}^{(i)})$, yielding $Q[\mathbf{D}^{(i)}] = \prod_{W^k \in \mathbf{D}^{(i)} \cap \mathbf{V}} \frac{Q[\mathbf{W}^{(k)}]}{Q[\mathbf{W}^{(k-1)}]}$. Since each $Q[\mathbf{W}^{(k)}]_\mathbf{s}$ is identifiable by induction, $\prod_{W^k \in \mathbf{D}^{(i)} \cap \mathbf{V}} \frac{Q[\mathbf{W}^{(k)}]_\mathbf{s}}{Q[\mathbf{W}^{(k-1)}]_\mathbf{s}}$ is also identifiable. □

Proof of Theorem 1. Suppose Algorithm 1 returns an identifying formula, there exists a CI-graph T under context $\mathbf{xy}$ such that the c-factors in the mutilated graph are deemed identifiable (lines 6-9). By Lemma 1, it suffices to show that a c-factor is indeed identifiable if ID-QFUNC returns an identifying formula for some context $\mathbf{s}$ and H in line 6.

Since $\mathbf{s}$ and H satisfy the condition in line 7, the c-factor $Q[\mathbf{C}_i]_\mathbf{xy}$ in T can be attained from the c-factor $Q[\mathbf{C}_i]_\mathbf{s}$ in H by selecting instantiations compatible with $\mathbf{xy}$. This is because the two c-factors involve the same set of CPTs, and the context $\mathbf{s}$ does not exclude any instantiations beyond those ruled out by the context $\mathbf{xy}$. Hence, $Q[\mathbf{C}_i]_\mathbf{xy}$ is identifiable if $Q[\mathbf{C}_i]_\mathbf{s}$ is identifiable. Since the identification method ID-QFUNC follows the two operations described in Lemma 2, the c-factor $Q[\mathbf{C}_i]_\mathbf{s}$ is identifiable if ID-QFUNC returns a formula for its estimation by the lemma. □

Proof of Proposition 3. When no CSI is imposed, i.e., $\mathcal{C} = \emptyset$, the CI-graph is identical to the original causal graph G. By setting $\mathbf{s} = \emptyset$ and $H = G$ in line 6, Algorithm 1 becomes identical to the IDENTIFY algorithm in [Tian and Pearl, 2003]. The operation of pruning leaves from $H_\mathbf{D}$ in lines 16-19 subsumes the operation of keeping the ancestors of $\mathbf{C}$ in IDENTIFY, and the operation of finding a maximal c-component in lines 20-22 is identical to the one in IDENTIFY. Since IDENTIFY has been shown to be sound and complete [Huang and Valtorta, 2006], Algorithm 1 inherits the same properties for classical identifiability. □

Proof of Theorem 2. Suppose there is a context-partition $\{\mathbf{s}_1, \ldots, \mathbf{s}_k\}$ over $\mathbf{V} \setminus (\mathbf{X} \cup \mathbf{Y})$ such that each $\Pr_{\mathbf{x}}(\mathbf{y}, \mathbf{s}_i)$ is identifiable by Algorithm 1. Then $\Pr_{\mathbf{x}}(\mathbf{y}) = \sum_{i=1}^{k} \Pr_{\mathbf{x}}(\mathbf{y}, \mathbf{s}_i)$ is identifiable by the law of total probability. $\square$

D.2 SEPARATION RESULTS

Proof of Proposition 1. The only-if direction directly follows from the definitions. We now show the contrapositive of the if direction. Suppose $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable, there exist some $\mathbf{x}'$ and $\mathbf{y}'$ such that $\Pr_{\mathbf{x}'}(\mathbf{y}')$ is unidentifiable. We can simply set $\mathbf{x} = \mathbf{x}'$ and $\mathbf{y} = \mathbf{y}'$ for the states in $\Pr_{\mathbf{x}}(\mathbf{y})$. $\square$

Proof of Proposition 2. If $\Pr_{\mathbf{x}}(\mathbf{y})$ is unidentifiable, then $\Pr_{\mathbf{X}}(\mathbf{Y})$ is also unidentifiable by the definition of variable-based identifiability. Now consider the opposite direction. If $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable, then there exist states $\mathbf{x}$ and $\mathbf{y}$ such that $\Pr_{\mathbf{x}}(\mathbf{y})$ is unidentifiable. We go over each state s in $\mathbf{xy}$ and create a new state s' (outside CSI contexts) for S. The CPTs for S and each of its children C are updated as follows:

$$\theta'_{s'|\mathbf{p}} = \theta_{s|\mathbf{p}} \cdot \epsilon, \quad \theta'_{s|\mathbf{p}} = \theta_{s|\mathbf{p}} \cdot (1-\epsilon), \quad \theta'_{c|\mathbf{p}_C, s'} = \theta_{c|\mathbf{p}_C, s}$$

where $\mathbf{P}_C$ denotes the parents of C (excluding S) and ϵ is a positive constant in range $(0, 1)$. The construction ensures that all CSIs satisfied by the original parameterization Θ are also satisfied by the new parameterization Θ'.

Let $\mathbf{x}'$ and $\mathbf{y}'$ be the results of replacing s by s' in $\mathbf{x}$ and $\mathbf{y}$. Then the update parameterization induces the following observational distribution $\Pr'(\mathbf{V})$ and causal effect $\Pr'_{\mathbf{x}'}(\mathbf{y}')$:

$$\Pr'(\mathbf{v}) = \begin{cases} \Pr(\mathbf{v}|_s) \cdot \epsilon & s' \in \mathbf{v} \\ \Pr(\mathbf{v}) \cdot (1-\epsilon) & s \in \mathbf{v} \\ \Pr(\mathbf{v}) & \text{otherwise} \end{cases}$$

$$\Pr'_{\mathbf{x}'}(\mathbf{y}') = \begin{cases} \Pr_{\mathbf{x}}(\mathbf{y}) \cdot \epsilon & S \in \mathbf{Y} \\ \Pr_{\mathbf{x}}(\mathbf{y}) & \text{otherwise} \end{cases}$$

where $\mathbf{v}|_s$ replaces the state of S in $\mathbf{v}$ by the state s. Hence, if there exist parameterizations Θ^1, Θ^2 that induce a same $\Pr(\mathbf{V})$ yet distinct $\Pr_{\mathbf{x}}(\mathbf{y})$, we can apply the above procedure to construct parameterizations that induce a same $\Pr'(\mathbf{V})$ yet distinct $\Pr_{\mathbf{x}'}(\mathbf{y}')$. By iterating over all states in $\mathbf{xy}$, we conclude that there exist $\mathbf{x}', \mathbf{y}'$ that do not involve any contexts of CSIs and for which $\Pr_{\mathbf{x}'}(\mathbf{y}')$ is unidentifiable. We finally set $\mathbf{x} = \mathbf{x}'$ and $\mathbf{y} = \mathbf{y}'$ for $\Pr_{\mathbf{x}}(\mathbf{y})$. $\square$

Proposition 6. *There exists a tuple $\langle G, \mathbf{V}, \mathcal{C}\rangle$ where $\mathcal{C}$ contains only CSI constraints for which $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable but $\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable for some states $\mathbf{x}, \mathbf{y}$.*

Proof. Consider the causal graph in Figure 1a. Let "entry-level" be state 0 of J, "$Y \geq 10$" be state 0 of Y, and "low" be state 0 of S. The CSI constraint can be written as $(S \perp\!\!\!\perp D | J = 0, Y)$.

We apply Algorithm 1 to show the identifiability of $\Pr_{Y=0}(J = 0, S = 0)$. Figure 3a depicts the CI-graph T under the context $\{Y = 0, J = 0, S = 0\}$ and Figure 3b depicts its mutilated graph. We are left to show that the c-factors $Q[ADJ]_{Y=0,J=0,S=0}$ and $Q[S]_{Y=0,J=0,S=0}$ in the mutilated graph are identifiable.

To identify the first c-factor, we consider the context $\mathbf{s} : \{J = 0\}$ and its corresponding CI-graph which coincides with T. Since the context and CI-graph satisfy the condition in line 7 of Algorithm 1, we can apply the ID-QFUNC procedure to establish the identifiability of $Q[ADJ]_{J=0}$. This is because the c-component $\{A, D, J\}$ can be obtained from CI-graph T by first finding the maximal c-component $\{A, D, J, Y\}$ and then pruning the leaf node Y. To identify the second c-factor, we again consider the context $\mathbf{s} : \{J = 0\}$ and its CI-graph depicted in T. Since $\{S\}$ is directly a maximal c-component in T, the c-factor $Q[S]_{J=0}$ is identifiable according to ID-QFUNC. Together, the causal effect $\Pr_{Y=0}(J = 0, S = 0)$ is identifiable since the entries in c-factors $Q[ADJ]_{Y=0,J=0,S=0}$ and $Q[S]_{Y=0,J=0,S=0}$ can be extracted from $Q[ADJ]_{J=0}$ and $Q[S]_{J=0}$.

We next show that the variable-based causal effect $\Pr_Y(J, S)$ is not identifiable. We assume that J has three states $\{0, 1, 2\}$ and all other variables are binary with states $\{0, 1\}$. For both parameterizations Θ_1, Θ_2, we assign uniform distributions for A, D, structural equation $Y = A$, and the following CPT for J:

$$\theta^1_{J|A,D} = \theta^2_{J|A,D} = \begin{cases} 0.01 & \text{if } J = 0 \\ 0.99 & \text{if } A \oplus D = J - 1 \\ 0 & \text{if } A \oplus D \neq J - 1 \end{cases}$$

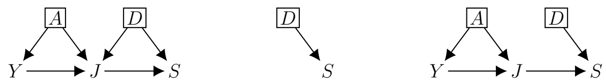
we then assign different CPTs for S:

$$\theta^1_{S|D,Y,J} = \begin{cases} 0.5 & \text{if } J = 0 \\ 0.99 & \text{if } J \neq 0 \text{ and } (J-1) \oplus D \oplus Y = S \\ 0.01 & \text{if } J \neq 0 \text{ and } (J-1) \oplus D \oplus Y \neq S \end{cases}$$

$$\theta^2_{S|D,Y,J} = \begin{cases} 0.5 & \text{if } J = 0 \\ 0.99 & \text{if } J \neq 0 \text{ and } S = 0 \\ 0.01 & \text{if } J \neq 0 \text{ and } S = 1 \end{cases}$$

The two parameterizations induce a same $\Pr(\mathbf{V})$ yet different $\Pr_{Y=0}(J = 1, S = 1)$. $\square$

Proposition 7. *There exists a tuple $\langle G, \mathbf{V}, \mathcal{C}\rangle$ where $\mathcal{C}$ contains only CSI constraints for which $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable wrt $\langle G, \mathbf{V}\rangle$ but is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}\rangle$.*



(a) CI-graph 1b (b) mutilated 9a (c) context $Y = 0$

Figure 9: context-induced and mutilated graphs for the examples in Figure 1 under the CSI $(J \perp\!\!\!\perp D \mid A, Y = 0)$.

Proof. Consider the causal graph in Figure 1b and the CSI $(J \perp\!\!\!\perp D \mid A, Y = 0)$. The classical causal effect $\Pr_J(S)$ is unidentifiable by the ID algorithm [Shpitser and Pearl, 2006] since the subgraph containing $\{A, Y, J\}$ forms a *hedge* wrt the causal effect.[12]

We next show that the causal effect is identifiable with the CSI constraint. By Proposition 2, the variable-based causal effect $\Pr_J(S)$ is identifiable iff the state-based causal effect $\Pr_{J=0}(S = 0)$ is identifiable since none of the contexts appear in the the CSI. The CI-graph under the context $\{J = 0, S = 0\}$ is identical to the original causal graph, and its mutilated graph T is depicted in Figure 9b. The only c-component in the mutilated graph is $\{S, D\}$ and its corresponding c-factor is $Q[SD]_{J=0,S=0}$. We now construct the context $\mathbf{s} : \{Y = 0\}$ with the corresponding CI-graph H shown in Figure 9c. Since $\mathbf{s}$ and H satisfy the condition in line 7 of Algorithm 1, $Q[SD]_{J=0,S=0}$ is identifiable if $Q[SD]_{Y=0}$ is identifiable by the ID-QFUNC procedure. This is indeed the case since the c-component $\{S, D\}$ is a maximal c-component in H. □

Before showing Proposition 8, we present the following Lemma 3 which will be used for the proof of Proposition 8.

Lemma 3. *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two sets of state constraints where $\mathcal{C}_1 \subseteq \mathcal{C}_2$.*[13] *Then a causal effect $\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1\rangle$ if it is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_2\rangle$.*

Proof. By induction, it suffices to show that an unidentifiable causal effect $\Pr_{\mathbf{x}}(\mathbf{y})$ remains unidentifiable if we add a new state to any variable W. Since $\Pr_{\mathbf{x}}(\mathbf{y})$ is unidentifiable, there exist two parameterizations Θ_1, Θ_2 that induce a same $\Pr(\mathbf{V})$ but different values for $\Pr_{\mathbf{x}}(\mathbf{y})$. Our proof is based on constructing parameterizations Θ'_1, Θ'_2 that contain an additional state for W while inducing unidentifiability.

Let $\mathbf{P} = \{P_1, \ldots, P_k\}$ and $\mathbf{C} = \{C_1, \ldots, C_t\}$ be the parents and children of W. Moreover, let $w_1, \ldots w_m$ be the current states and w' be the new state of W. We construct new parameterizations Θ'_1, Θ'_2 from the original parameterizations Θ_1, Θ_2 as follows. Let ϵ be a positive constant in $(0, 1)$. We modify the CPTs for W and each child $C \in \mathbf{C}$:

$$\theta'_{w'|\mathbf{p}} = \theta_{w_1|\mathbf{p}} \cdot \epsilon \quad \theta'_{w_1|\mathbf{p}} = \theta_{w_1|\mathbf{p}} \cdot (1 - \epsilon)$$

$$\theta'_{c|\mathbf{p}_C,w'} = \theta_{c|\mathbf{p}_C,w_1}$$

where $\mathbf{P}_C$ denotes the parents of C excluding W. Let $\Pr'_1$ and $\Pr'_2$ be the distribution induced by Θ'_1 and Θ'_2. For each instantiation $(\mathbf{u}, \mathbf{v})$ over hidden and observed variables,

$$\Pr'(\mathbf{u}, \mathbf{v}) = \begin{cases} \Pr((\mathbf{u}, \mathbf{v})|_{w_1}) \cdot \epsilon & w' \in (\mathbf{u}, \mathbf{v}) \\ \Pr(\mathbf{u}, \mathbf{v}) \cdot (1 - \epsilon) & w_1 \in (\mathbf{u}, \mathbf{v}) \\ \Pr(\mathbf{u}, \mathbf{v}) & \text{otherwise} \end{cases}$$

where $(\mathbf{u}, \mathbf{v})|_{w_1}$ replaces the state of W in $(\mathbf{u}, \mathbf{v})$ by w_1. Suppose W is hidden, summing out $\mathbf{U}$ yields $\Pr'(\mathbf{V}) = \Pr(\mathbf{V})$; hence, $\Pr'_1(\mathbf{V}) = \Pr_1(\mathbf{V}) = \Pr_2(\mathbf{V}) = \Pr'_2(\mathbf{V})$. Suppose W is observed, we can compute $\Pr'(\mathbf{V})$ as follows:

$$\Pr'(\mathbf{v}) = \begin{cases} \Pr(\mathbf{v}|_{w_1}) \cdot \epsilon & w' \in \mathbf{v} \\ \Pr(\mathbf{v}) \cdot (1 - \epsilon) & w_1 \in \mathbf{v} \\ \Pr(\mathbf{v}) & \text{otherwise} \end{cases}$$

Again, $\Pr'_1(\mathbf{V}) = \Pr'_2(\mathbf{V})$ since $\Pr_1(\mathbf{V}) = \Pr_2(\mathbf{V})$.

We are left to show that $\Pr'_{1\mathbf{x}}(\mathbf{y}) \neq \Pr'_{2\mathbf{x}}(\mathbf{y})$. Suppose $W \in \mathbf{U}$, the distribution $\Pr_{\mathbf{x}}(\mathbf{V})$ is preserved and therefore $\Pr'_{1\mathbf{x}}(\mathbf{y}) = \Pr_{1\mathbf{x}}(\mathbf{y}) \neq \Pr_{2\mathbf{x}}(\mathbf{y}) = \Pr'_{2\mathbf{x}}(\mathbf{y})$. Suppose $W \in \mathbf{V}$, we can again compute $\Pr'_{\mathbf{x}}(\mathbf{y})$ as follows:

$$\Pr'_{\mathbf{x}}(\mathbf{y}) = \begin{cases} \Pr_{\mathbf{x}}(\mathbf{y}) & W \notin \mathbf{Y} \\ \Pr_{\mathbf{x}}(\mathbf{y}|_{w_1}) \cdot \epsilon & W \in \mathbf{Y} \text{ and } w' \in \mathbf{y} \\ \Pr_{\mathbf{x}}(\mathbf{y}) \cdot (1 - \epsilon) & W \in \mathbf{Y} \text{ and } w_1 \in \mathbf{y} \end{cases}$$

In all cases, $\Pr'_{1\mathbf{x}}(\mathbf{y}) \neq \Pr'_{2\mathbf{x}}(\mathbf{y})$. □

Proposition 8. *Let $\mathcal{C}$ be a set of state constraints and G be a Semi-Markovian graph. A causal effect $\Pr_{\mathbf{X}}(\mathbf{Y})$ (or $\Pr_{\mathbf{x}}(\mathbf{y})$) is identifiable wrt $\langle G, \mathbf{V}\rangle$ iff $\Pr_{\mathbf{X}}(\mathbf{Y})$ (or $\Pr_{\mathbf{x}}(\mathbf{y})$) is constrained-identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}\rangle$.*

Proof. The only-if direction follows from the definition. We now prove the contrapositive of the if direction. According to the ID algorithm [Shpitser and Pearl, 2006], $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable iff there exist two parameterizations Θ^1 and Θ^2 that induce a same $\Pr(\mathbf{V})$ yet different $\Pr_{\mathbf{x}}(\mathbf{y})$ and in which *all variables are binary*. Hence, we can apply Lemma 3 to construct parameterizations with arbitrary numbers of variable states that induce unidentifiability. The equivalence between $\Pr_{\mathbf{X}}(\mathbf{Y})$ and $\Pr_{\mathbf{x}}(\mathbf{y})$ follows from the fact that we can always permute the states of $\mathbf{X}$ and $\mathbf{Y}$. □

Proposition 9. *There exist a causal graph G, observed variables $\mathbf{V}$, CSI constraints $\mathcal{C}_1$, and state constraints $\mathcal{C}_2$ for which $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1\rangle$ but is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1 \cup \mathcal{C}_2\rangle$.*

Proof. Consider the causal graph in Figure 2a and the CSI constraints $(X \perp\!\!\!\perp U|A = 0)$ and $(Y \perp\!\!\!\perp U|X, A = 1)$. We

[12] A *hedge* is a graphical notion that characterizes the unidentifiability of causal effects; see [Shpitser and Pearl, 2006] for details.

[13] We say $\mathcal{C}_1 \subseteq \mathcal{C}_2$ if for each variable T, the states of T in $\mathcal{C}_1$ are a subset of the states of T in $\mathcal{C}_2$.

first show that the variable-based causal effect is unidentifiable without restricting variable states. Assume variables U, X, Y have states $\{0, 1\}$ and A has states $\{0, 1, 2\}$. We construct two parameterizations Θ^1, Θ^2 that induce a same $\Pr(\mathbf{V})$ yet distinct answers for the causal effect $\Pr_{X=0}(Y = 0)$. In both parameterizations, we assign uniform distributions for U, A and the following CPT for X :

$$\theta^1_{X|U,A} = \theta^2_{X|U,A} = \begin{cases} 0.5 & \text{if } A = 0 \text{ or } A = 1 \\ 1 & \text{if } A = 2 \text{ and } X = U \\ 0 & \text{if } A = 2 \text{ and } X \neq U \end{cases}$$

We then assign different CPTs for Y as follows:

$$\theta^1_{Y|U,X,A} = \begin{cases} 0.5 & \text{if } A = 0 \text{ or } A = 1 \\ 0.99 & \text{if } A = 2 \text{ and } U \oplus X = Y \\ 0.01 & \text{if } A = 2 \text{ and } U \oplus X \neq Y \end{cases}$$

$$\theta^2_{Y|U,X,A} = \begin{cases} 0.5 & \text{if } A = 0 \text{ or } A = 1 \\ 0.99 & \text{if } A = 2 \text{ and } Y = 0 \\ 0.01 & \text{if } A = 2 \text{ and } Y = 1 \end{cases}$$

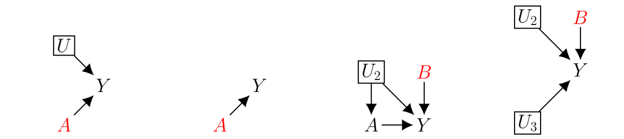
The two parameterizations induce a same $\Pr(\mathbf{V})$ yet distinct $\Pr_{X=0}(Y)$ and $\Pr_{X=1}(Y)$.

We next apply Algorithm 2 to show that the causal effect $\Pr_X(Y)$ is identifiable if A has states $\{0, 1\}$. We consider the context-partition $\{A{=}0, A{=}1\}$ and reduce the problem of identifying $\Pr_X(Y)$ to that of identifying sub-queries $\Pr_{X=0}(Y = 0, A = 0)$ and $\Pr_{X=0}(Y = 0, A = 1)$. Figures 5a and 5b depict the CI-graphs for the two sub-queries, and Figures 10a and 10b depict their mutilated graphs.

To identify the two c-factors in Figures 10a, we consider the context $A{=}0$ and its CI-graph T which coincides with Figure 5a. Since $\{U, Y\}$ and $\{A\}$ are maximal c-components in T, their c-factors $Q[UY]_{A=0}$ and $Q[A]_{A=0}$ are identifiable. It follows that $Q[UY]_{A=0,X=0,Y=0}$ and $Q[A]_{A=0,X=0,Y=0}$ are also identifiable, which concludes the identifiability of $\Pr_{X=0}(Y = 0, A = 0)$.

To identify the two c-factors in Figures 10b, we consider the context $A{=}1$ and its CI-graph T which coincides with Figure 5b. Since both $\{Y\}$ and $\{A\}$ are c-components in T, their corresponding c-factors $Q[Y]_{A=1}$ and $Q[A]_{A=1}$ are computable from $\Pr(\mathbf{V})$, which establishes the identifiability of $\Pr_{X=0}(Y = 0, A = 1)$. $\square$

Proposition 10. *There exist a causal graph G, observed variables $\mathbf{V}$, CSI constraints $\mathcal{C}_1$, state constraints $\mathcal{C}_2$, and particular states $\mathbf{x}$, and $\mathbf{y}$ where $\Pr_{\mathbf{x}}(\mathbf{y})$ is unidentifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1\rangle$ but is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1 \cup \mathcal{C}_2\rangle$. Moreover, $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1 \cup \mathcal{C}_2\rangle$.*



(a) mutilated 5a (b) mutilated 5b (c) mutilated 5c (d) mutilated 5d

Figure 10: mutilated graphs for the examples in Figure 5.

Proof. Consider the second example for Figure 2b in Section 4. For simplicity, let *raining* be the state 0 and *snowing* be the state 1 of B, and let *low-cost* be the state 0 of X. We can rewrite the CSIs as $(A \perp\!\!\!\perp U_1 \mid X = 0, U_2)$, $(Y \perp\!\!\!\perp U_3 \mid B = 0, U_2, A)$, and $(Y \perp\!\!\!\perp A | B = 1, U_2, U_3)$.

We first show $\Pr_{X=0}(Y = 0)$ is unidentifiable by finding two parameterizations Θ^1, Θ^2 that induce a same $\Pr(\mathbf{V})$ yet different $\Pr_{X=0}(Y = 0)$. We assume that all variables have states $\{0, 1\}$, except for B, which has states $\{0, 1, 2\}$. For both parameterizations, we assign uniform distributions for U_1, U_2, U_3, structural equations $X = U_3$, $A = U_2 \oplus X$, and the following CPT for B:

$$\theta^1_B = \theta^2_B = \begin{cases} 0.05 & \text{if } B = 0 \text{ or } B = 1 \\ 0.9 & \text{if } B = 2 \end{cases}$$

We assign different CPTs for Y in Θ_1 and Θ_2:

$$\theta^1_{Y|U_2,U_3,A,B} = \begin{cases} 0.5 & \text{if } B = 0 \text{ or } B = 1 \\ 0.99 & \text{if } B = 1 \text{ and } Y = U_2 \oplus A \oplus U_3 \\ 0.01 & \text{if } B = 1 \text{ and } Y \neq U_2 \oplus A \oplus U_3 \end{cases}$$

$$\theta^2_{Y|U_2,U_3,A,B} = \begin{cases} 0.5 & \text{if } B = 0 \text{ or } B = 1 \\ 0.99 & \text{if } B = 1 \text{ and } Y = 0 \\ 0.01 & \text{if } B = 1 \text{ and } Y = 1 \end{cases}$$

The two parameterizations attains a same $\Pr(\mathbf{V})$ yet different answers for $\Pr_{X=0}(Y = 0)$.

We now apply Algorithm 2 to show that $\Pr_{X=0}(Y = 0)$ is identifiable if the states of B are restricted to $\{0, 1\}$. We consider the context-partition $\{B{=}0, B{=}1\}$ and show that the sub-queries $\Pr_{X=0}(Y = 0, B = 0)$ and $\Pr_{X=0}(Y = 0, B = 1)$ are identifiable. Figures 5c and 5d depict the CI-graphs for the sub-queries, and Figures 10c and 10d depict their mutilated graphs.

We first identify the c-factors of c-components $\{U_2, A, Y\}$ and $\{B\}$ in the mutilated graph for sub-query $\Pr_{X=0}(Y = 0, B = 0)$. Consider the context $\mathbf{s} : \{X{=}0, B{=}0\}$ and its CI-graph H which coincides with Figure 10c (Algorithm 1 lines 6-7). Since both $\{U_2, A, Y\}$ and $\{B\}$ are maximal c-components in H, $\Pr_{X=0}(Y = 0, B = 0)$ is identifiable.

We now identify the c-factors of c-components $\{U_2, U_3, Y\}$ and $\{B\}$ in the mutilate graph for sub-query $\Pr_{X=0}(Y = 0, B = 1)$. We consider the context $\mathbf{s} : \{B{=}1\}$ and its CI-graph H shown in Figure 11. The first c-component can be

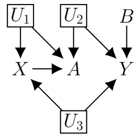


Figure 11: CI-graph of Figure 2b under the context B=1.

obtained from H by first pruning leaves A, X and then taking the maximal c-component. The second c-component is a maximal c-component of H. Hence, $\Pr_{X=0}(Y=0, B=1)$ is identifiable by the soundness of Algorithm 1.

We finally show that $\Pr_X(Y)$ is unidentifiable under both the CSI and state constraints by constructing parameterizations Θ^1, Θ^2 that form an instance of unidentifiability. We assume all variables are binary with states $\{0,1\}$, except for X, which has states $\{0,1,2\}$. For both Θ^1 and Θ^2, we assign uniform distributions for variables U_1, U_2, U_3, B and the following CPTs for X, Y:

$$\theta^1_{X|U_1,U_3} = \theta^2_{X|U_1,U_3} = \begin{cases} (0.05, 0.95, 0) & \text{if } U_1 = 0 \\ (0.05, 0, 0.95) & \text{if } U_1 = 1 \end{cases}$$

$$\theta^1_{Y|U_2,U_3,A,B} = \theta^2_{Y|U_2,U_3,A,B} = \begin{cases} 0.99 & \text{if } A = Y \\ 0.01 & \text{if } A \neq Y \end{cases}$$

We assign distinct CPTs for A in Θ^1 and Θ^2:

$$\theta^1_{A|U_1,U_2,X} = \begin{cases} 0.5 & \text{if } X = 0 \\ 0.99 & \text{if } X \neq 0 \text{ and } (X-1) \oplus U_1 = A \\ 0.01 & \text{if } X \neq 0 \text{ and } (X-1) \oplus U_1 \neq A \end{cases}$$

$$\theta^2_{A|U_1,U_2,X} = \begin{cases} 0.5 & \text{if } X = 0 \\ 0.99 & \text{if } X \neq 0 \text{ and } A = 0 \\ 0.01 & \text{if } X \neq 0 \text{ and } A = 1 \end{cases}$$

The parameterizations induce a same $\Pr(\mathbf{V})$ yet different values for $\Pr_{X=1}(Y=1)$. $\square$

Proposition 11. *Let $\mathcal{C}_1$ be a set of CSI constraints and $\mathcal{C}_2$ be a set of state constraints. Suppose $\mathcal{C}_2$ mentions treatment variables $\mathbf{X}_1$ and outcome variables $\mathbf{Y}_1$. Let $\mathbf{X}_2 = \mathbf{X} \setminus \mathbf{X}_1$ and $\mathbf{Y}_2 = \mathbf{Y} \setminus \mathbf{Y}_1$. Then $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable (wrt $\langle G, \mathbf{V}, \mathcal{C}_1 \cup \mathcal{C}_2 \rangle$) iff $\Pr_{\mathbf{x}_1\mathbf{x}_2}(\mathbf{y}_1\mathbf{y}_2)$ is identifiable for all assignments $\mathbf{x}_1, \mathbf{y}_1$ specified by $\mathcal{C}_2$ and any assignment $\mathbf{x}_2, \mathbf{y}_2$ that do not appear in the contexts of $\mathcal{C}_1$.*

Proof. The only-if direction follows from the definition of variable-based identifiability (Definition 3). We now prove the contrapositive of the if direction. Suppose $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable, there exist two parameterizations that induce a same $\Pr(\mathbf{V})$ but different $\Pr_{\mathbf{x}_1\mathbf{x}_2}(\mathbf{y}_1\mathbf{y}_2)$, where $\mathbf{x}_1, \mathbf{y}_1$ are states specified by $\mathcal{C}_2$. We then iterate over each state s in $\mathbf{x}_2\mathbf{y}_2$ and apply the construction method from the proof of Proposition 2. The resulting parameterizations satisfy $\mathcal{C}_1$ and $\mathcal{C}_2$, induce a same $\Pr(\mathbf{V})$, and yield different values of $\Pr_{\mathbf{x}_1\mathbf{x}'_2}(\mathbf{y}_1\mathbf{y}'_2)$ for $\mathbf{x}'_2, \mathbf{y}'_2$ outside the CSI contexts. $\square$

Corollary 1. *Let $\mathcal{C}_1$ be a set of CSI constraints and $\mathcal{C}_2$ be a set of state constraints. If the states $\mathbf{x}, \mathbf{y}$ do not appear in the context of $\mathcal{C}_1$ and variables $\mathbf{X}, \mathbf{Y}$ are not mentioned in $\mathcal{C}_2$, then $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1 \cup \mathcal{C}_2 \rangle$ iff $\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable wrt $\langle G, \mathbf{V}, \mathcal{C}_1 \cup \mathcal{C}_2 \rangle$.*

Proof. Since $\mathcal{C}_2$ does not mention variables $\mathbf{X}$ and $\mathbf{Y}$, the variable sets in Proposition 11 take the following values: $\mathbf{X}_1 = \emptyset$, $\mathbf{Y}_1 = \emptyset$, $\mathbf{X}_2 = \mathbf{X}$, and $\mathbf{Y}_2 = \mathbf{Y}$. Hence, $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable iff $\Pr_{\mathbf{x}}(\mathbf{y})$ is identifiable for any $\mathbf{x}, \mathbf{y}$ not appearing in the CSI contexts. $\square$

n	r	stats	CLASSICAL-ID	ID-CSI-HEU		ID-CSI-STATE-HEU		DOSEARCH
			variable & state	variable	state	variable	state	variable
6	0.2	#ID/success	171/200	171/200	171/200	171/200	171/200	171/200
		runtime (s)	0.001	0.001	0.001	0.002	0.002	0.9
	0.5	#ID/success	171/200	172/200	174/200	173/200	176/200	173/200
		runtime (s)	0.001	0.001	0.002	0.002	0.002	0.9
	0.8	#ID/success	171/200	171/200	172/200	175/200	176/200	174/200
		runtime (s)	0.001	0.001	0.001	0.002	0.002	0.9
8	0.2	#ID/success	120/200	120/200	123/200	121/200	124/200	120/200
		runtime (s)	0.002	0.002	0.003	0.003	0.002	6.2
	0.5	#ID/success	120/200	121/200	132/200	123/200	133/200	122/199
		runtime (s)	0.002	0.002	0.002	0.005	0.003	4.9
	0.8	#ID/success	120/200	121/200	133/200	126/200	137/200	123/198
		runtime (s)	0.002	0.002	0.002	0.005	0.003	6.6
10	0.2	#ID/success	115/200	115/200	118/200	115/200	118/200	112/198
		runtime (s)	0.003	0.004	0.003	0.004	0.003	14.0
	0.5	#ID/success	115/200	117/200	127/200	120/200	132/200	116/191
		runtime (s)	0.002	0.003	0.003	0.005	0.005	42.5
	0.8	#ID/success	115/200	119/200	132/200	129/200	140/200	124/190
		runtime (s)	0.002	0.003	0.003	0.007	0.004	40.2
12	0.2	#ID/success	107/200	108/200	113/200	109/200	114/200	99/175
		runtime (s)	0.003	0.003	0.004	0.005	0.004	92.2
	0.5	#ID/success	107/200	109/200	121/200	114/200	124/200	79/122
		runtime (s)	0.003	0.003	0.003	0.007	0.005	165.3
	0.8	#ID/success	107/200	109/200	131/200	121/200	136/200	79/127
		runtime (s)	0.003	0.003	0.003	0.009	0.005	152.0
15	0.2	#ID/success	84/200	84/200	90/200	85/200	90/200	-
		runtime (s)	0.004	0.005	0.004	0.007	0.006	-
	0.5	#ID/success	84/200	84/200	101/200	87/200	104/200	-
		runtime (s)	0.004	0.004	0.005	0.01	0.01	-
	0.8	#ID/success	84/200	85/200	108/200	89/200	110/200	-
		runtime (s)	0.004	0.004	0.004	0.02	0.01	-
20	0.2	#ID/success	69/200	69/200	72/200	70/200	72/200	-
		runtime (s)	0.007	0.007	0.007	0.01	0.008	-
	0.5	#ID/success	69/200	72/200	86/200	78/200	90/200	-
		runtime (s)	0.007	0.007	0.007	0.03	0.01	-
	0.8	#ID/success	69/200	76/200	95/200	84/200	100/200	-
		runtime (s)	0.003	0.003	0.003	0.14	0.02	-
30	0.2	#ID/success	45/200	45/200	47/200	45/200	47/200	-
		runtime (s)	0.03	0.03	0.03	0.05	0.03	-
	0.5	#ID/success	45/200	48/200	61/200	51/200	62/200	-
		runtime (s)	0.03	0.03	0.03	0.17	0.05	-
	0.8	#ID/success	45/200	48/200	65/200	51/200	69/200	-
		runtime (s)	0.03	0.03	0.03	0.31	0.07	-
40	0.2	#ID/success	20/200	21/200	24/200	22/200	25/200	-
		runtime (s)	0.03	0.03	0.03	0.06	0.04	-
	0.5	#ID/success	20/200	22/200	30/200	22/200	30/200	-
		runtime (s)	0.03	0.03	0.03	0.19	0.06	-
	0.8	#ID/success	20/200	22/200	41/200	28/200	44/200	-
		runtime (s)	0.03	0.03	0.03	0.70	0.09	-
50	0.2	#ID/success	15/200	15/200	19/200	15/200	19/200	-
		runtime (s)	0.06	0.06	0.06	0.24	0.08	-
	0.5	#ID/success	15/200	19/200	31/200	19/199	33/200	-
		runtime (s)	0.06	0.07	0.06	2.68	0.21	-
	0.8	#ID/success	15/200	20/200	37/200	25/200	41/200	-
		runtime (s)	0.06	0.07	0.06	4.08	0.37	-

Table 1: We record the number of variable-based (column "variable") and state-based (column "state") causal effects deemed identifiable for each method under causal graphs with varying numbers of nodes (n) and CSIs (r). We also record the number of runs that complete within the 5-minute timeout (labeled as "success" in the table). For runs that exceed the timeout, we set the runtime to 5 minutes when computing the average.