Article

Universal Causality

Sridhar Mahadevan

Adobe Research, 345 Park Avenue, San Jose, CA 95110, USA; smahadev@adobe.com

Abstract: Universal Causality is a mathematical framework based on higher-order category theory, which generalizes previous approaches based on directed graphs and regular categories. We present a hierarchical framework called UCLA (Universal Causality Layered Architecture), where at the top-most level, causal interventions are modeled as a higher-order category over simplicial sets and objects. Simplicial sets are contravariant functors from the category of ordinal numbers Δ into sets, and whose morphisms are order-preserving injections and surjections over finite ordered sets. Non-random interventions on causal structures are modeled as face operators that map n-simplices into lower-level simplices. At the second layer, causal models are defined as a category, for example defining the schema of a relational causal model or a symmetric monoidal category representation of DAG models. The third layer corresponds to the data layer in causal inference, where each causal object is mapped functorially into a set of instances using the category of sets and functions between sets. The fourth homotopy layer defines ways of abstractly characterizing causal models in terms of homotopy colimits, defined in terms of the nerve of a category, a functor that converts a causal (category) model into a simplicial object. Each functor between layers is characterized by a universal arrow, which define universal elements and representations through the Yoneda Lemma, and induces a Grothendieck category of elements that enables combining formal causal models with data instances, and is related to the notion of *ground graphs* in relational causal models. Causal inference between layers is defined as a lifting problem, a commutative diagram whose objects are categories, and whose morphisms are functors that are characterized as different types of fibrations. We illustrate UCLA using a variety of representations, including causal relational models, symmetric monoidal categorical variants of DAG models, and non-graphical representations, such as integer-valued multisets and separoids, and measure-theoretic and topological models.

Keywords: artificial intelligence; higher-order category theory; causality; machine learning; statistics

1. Introduction

Applied category theory [1] has been used to design algorithms for dimensionality reduction and data visualization [2], resolve impossibility theorems in data clustering [3] and propose schemes for knowledge representation [4]. Universal Causality (UC) is a mathematical framework based on applied higher-order category theory, which applies to graph-based [5] and non-graphical representations [6–8], and statistical [9] and non-statistical frameworks [10,11] (see Table 1 and Figure 1). Ordinary categories are defined as a collection of *objects* that interact pairwise through a collection of *morphisms*. Higher-order categories, such as simplicial sets [12], quasicategories [13] and ∞-categories [14], model higher-order interactions among groups of objects, and generalize both directed graphs and ordinary categories. Our approach builds extensively on categories over *functors*. Causal interventions are defined over the functor category of simplicial objects, mapping ordinal numbers into sets or category objects. Causal inference is defined over the functor category of presheafs $\mathbf{Hom}_{\mathcal{C}}(-, c)$, mapping an object c in category $\mathcal{C}$ into the set of morphisms into it. *Adjoint functors* define a pair of opposing functors between categories. Causal models are often characterized in terms of their underlying conditional independence structures. We model this relationship by adjoint functors between the category of conditional independence structures [15], based on algebraic representations such as *separoids* [10], and

the category of causal models, defined by graphical approaches [16] or non-graphical approaches, such as integer-valued multisets [8] or measure-theoretic information fields [6,7]. We build extensively on *universal constructions*, such as colimits and limits, defined through *lifting diagrams* [17].

Table 1. Category theory provides a unifying mathematical framework for relating the diverse formalisms used to study causal inference.

| Representation	Objects	Morphisms	Citation
Rank-ordered statistics	Plants	Total ordering	Darwin [18]
Structural equation models	Variables	Algebraic equations	Wright [19]
Potential outcomes	Humans	Drug effects	Imbens and Rubin [9]
Directed Acyclic Graphs	Vertices	Paths	Pearl [5]
Distributive lattices	Subsets	Joins/Meets	Beerenwinkel et al. [20]
Relational causal models	Database schemas	Database relations	Maier et al. [21]
Information fields	Measurable Spaces	Measurable functions	Witsenhausen [6]
Resource Models	Monoidal resources	Profunctors	Fong and Spivak [1]
Universal Decision Models	UDM States	UDM morphisms	Mahadevan [22]
Counterfactual logic	Propositions	Proofs	Lewis [11]
Variational inequalities	Consumers/Producers	Trade	Nagurney [23]
Discourse sheaves	Users	Communication	Hansen and Ghrist [24]
String diagram surgery	Tensored objects	Tensored morphisms	Jacobs et al. [25]
Mean embeddings	RKHS embeddings	Mean maps	Muandet et al. [26]
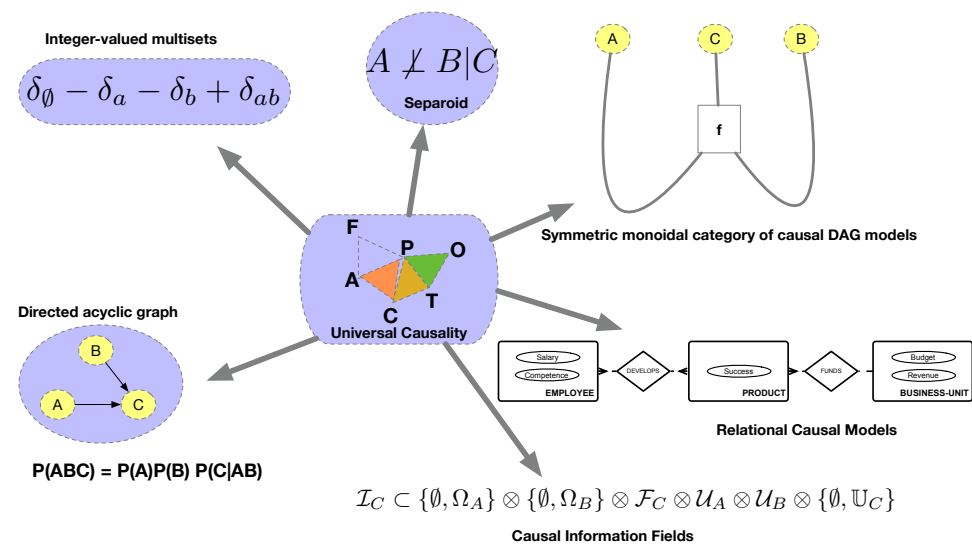
Figure 1. UC is a *representation-independent* framework that can be applied to many causal representations.

Over the past 150 years, causality has been studied using diverse formalisms (Table 1). While causal effects are inherently *directional*, differing from symmetric statistical models of correlation and invertible Bayesian inference, many causal discovery methods rely on querying a (symmetric) conditional independence oracle on submodels resulting from interventions on arbitrary subsets of variables (such as a *separating set* [27,28]). Abstractly, we can classify the causal representations in Table 1 using category theory in terms of their underlying *objects* and their associated *morphisms*. Causal morphisms can be algebraic, graph-based, logical, measure-theoretical, probabilistic or topological. For example, counterfactual mean embeddings [26] generalizes Rubin's potential outcome model to reproducing kernel Hilbert spaces (RKHS), where the kernel mean map is used to embed

a distribution in an RKHS, and the average treatment effect (ATE) is computed using mean maps. As Figure 1 emphasizes, UC is representation agnostic, and while it is related to category-theoretic approaches of causal DAG models that use symmetric monoidal categories [25,29,30], it differs substantially in many ways. UC introduces many novel ideas into the study of causal inference, including higher-order categorical structures based on simplicial sets and objects [12–14,31], adjoint functors mapping categories based on algebraic models of conditional independence [10] into actual causal models, lifting diagrams [17], and Grothendieck's category of elements that generalizes the notion of *ground graphs* in relational causal models [32]. As we show later, any category, including symmetric monoidal categories, can be converted into simplicial objects by using the nerve functor, but its left adjoint that maps a simplicial set into a category is lossy, and preserves structure only up to $n \leq 2$-simplices. Higher-order category structures can be useful in modeing causal inference under interference [33], where the traditional stable unit treatment value assumption (SUTVA) is violated. Higher-order categories can also help model hierarchical interventions over groups of objects.

As Studeny [8] points out, Bayesian DAG models capture only a small percentage of all conditional independence structures. In particular, the space of DAG models grows exponentially in the number of variables, whereas the number of conditional independence structures grows super-exponentially proportional to the number of Boolean functions. Consequently, UC is intended to be a general framework that applies to representations that are more expressive than DAG models. In particular, UC can be used to analyze recent work on relational causal models [21,32]. The notion of a *ground graph* in relational causal models is a special case of the Grothendieck category of elements that plays a key role in the UCLA architecture. UC applies equally well to non-graphical algebraic representations that are much more expressive than DAG models, including integer-valued multisets [8], separoids [10], as well as measure-theoretic representations, such as causal information fields [6,7], that have been shown to generalize Pearl's d-separation calculus [5].

Specifically, taking the simple example of a collider in Figure 1, in the Bayesian DAG parameterization, a well-established theoretical framework [34] specifies how to decompose the overall probability distribution into a product of local distributions. In contrast, in causal information fields [6,7], each variable is defined as a measurable space over a discrete or continuous set, and each local function is defined as a measurable function over its information field. For example, the information field $\mathcal{I}_C$ for variable C is defined to be some measurable subset over a product σ-algebra that includes the σ-algebras $\mathcal{U}_A$ and $\mathcal{U}_B$ over its parents A and B, but the information field of C cannot depend on its own values, hence its local σ-algebra is defined as $\{\emptyset, \mathbb{U}_C\}$, where $\mathbb{U}_C$ is the space of possible values of C. A full discussion of causal information fields is given in [7], who show it generalizes d-separation to models that include cycles and other more complex structures. Similarly, Studeny [8] proposed an algebraic framework called integer-valued multisets (imsets) for representing conditional independence structures far more expressive than DAG models. For the specific case of a DAG model $G = (V, E)$, an imset in standard form [8] is defined as

$$u_G = \delta_V - \delta_\emptyset + \sum_{i \in V} (\delta_{\mathbf{Pa}_i} - \delta_{i \cup \mathbf{Pa}_i})$$

where each δ_V term is the characteristic function associated with a set of variables V. Finally, separoids [10] is an algebraic framework for characterizing conditional independence as an abstract property, defined by a join semi-lattice equipped with a partial ordering $\leq$, and a ternary property $\perp\!\!\!\perp$ over triples of elements such that $X \perp\!\!\!\perp Y|Z$ defines the property that X is conditionally independent of Y given Z. It is worth pointing out that separoids are more general than the graphoid axiomatization [16] that underpins causal DAG models, since as Studeny [8] shows, graphoids are defined in terms of disjoint subsets of variables, which seriously limits their expressiveness. All these non-graphical representations can be naturally modeled within the UC framework. One of the unique aspects of UC is that causal interventions are themselves modeled as a (higher-order) category. Many approaches to

causal discovery use a sequence of interventions, which naturally compose and form a category. To achieve representation independence, we model interventions as a higher-order category defined by simplicial sets and objects [12]. One strength of the simplicial objects framework for modeling causal interventions is that it enables modeling hierarchical interventions over groups of objects.

UC builds on the concept of *universal arrows* [35] to illuminate in a representation-independent manner the central abstractions employed in causal inference. Figure 2 explains this concept with an example, which also illustrates the connection between categories and graphs. For every (directed) graph G, there is a universal arrow from G to the "forgetful" functor U mapping the category **Cat** of all categories to **Graph**, the category of all (directed) graphs, where for any category C, its associated graph is defined by $U(C)$. Intuitively, this forgetful functor "throws" away all categorical information, obliterating for example the distinction between the primitive morphisms f and g vs. their compositions $g \circ f$, both of which are simply viewed as edges in the graph $U(C)$. To understand this functor, consider a directed graph $U(C)$ defined from a category C, forgetting the rule for composition. That is, from the category C, which associates to each pair of composable arrows f and g, the composed arrow $g \circ f$, we derive the underlying graph $U(C)$ simply by forgetting which edges correspond to elementary arrows, such as f or g, and which are composites. For example, consider a partial order as the category $\mathcal{C}$, and then define $U(\mathcal{C})$ as the directed graph that results from the transitive closure of the partial ordering.

The universal arrow from a graph G to the forgetful functor U is defined as a pair $\langle C, u : G \to U(C) \rangle$, where u is a "universal" graph homomorphism. This arrow possesses the following *universal property*: for every other pair $\langle D, v : G \to H \rangle$, where D is a category, and v is an arbitrary graph homomorphism, there is a functor $f' : C \to D$, which is an arrow in the category **Cat** of all categories, such that *every* graph homomorphism $\phi : G \to H$ uniquely factors through the universal graph homomorphism $u : G \to U(C)$ as the solution to the equation $\phi = U(f') \circ u$, where $U(f') : U(C) \to H$ (that is, $H = U(D)$). Namely, the dotted arrow defines a graph homomorphism $U(f')$ that makes the triangle diagram "commute", and the associated "extension" problem of finding this new graph homomorphism $U(f')$ is solved by "lifting" the associated category arrow $f' : C \to D$. This property of universal arrows, as we show in the paper, provide the conceptual underpinnings of universal causality in the UCLA architecture, leading to the defining property of a universal causal representation through the Yoneda Lemma [35]. Recent work on causal discovery of DAG models [27,28] can be seen as restricted ways of defining adjoint functors between causal categories of DAG models and their underlying graphs, assuming access to a conditional independence oracle that can be queried on causal sub-models resulting from interventions on arbitrary subsets of variables.

Universal causal models are defined in terms of universal constructions, such as the pullback, pushforward, (co)equalizer, and (co)limits. Figure 3 illustrates how universal causal models are functors that map from some indexing category of abstract diagrams into an actual causal model. For instance, **COVID-19 Lockdown** caused a reduction in **Traffic** and **Agricultural Fires**, which in turn caused a significant reduction in **Pollution**. In UC, we are interested in a deeper question, namely whether the pullback of **Traffic** and **Agricultural Fires** could have been some other common cause that mediated between **COVID-19 Lockdown** and its effects. If such a common cause exists, it will be viewed as a limit of an abstract causal diagram, a functor that maps from the indexing category of all diagrams to the actual causal model shown.
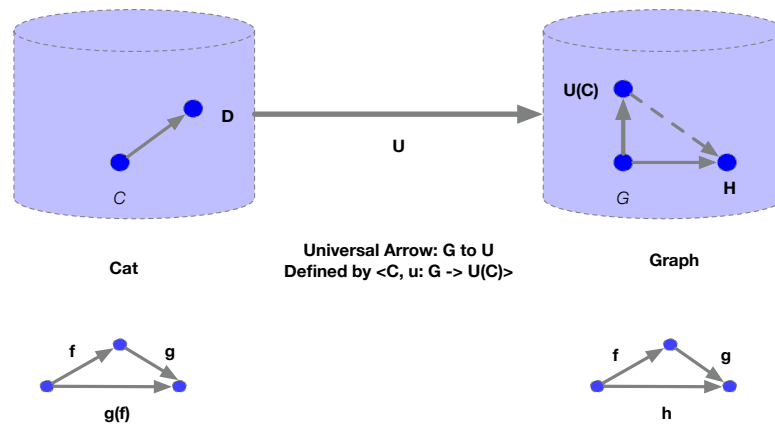
Figure 2. Universal arrows play a central role in the UCLA framework. In this example, the forgetful functor U between **Cat**, the category of all categories, and **Graph**, the category of all (directed) graphs, maps any category into its underlying graph, forgetting which arrows are primitive and which are compositional. The universal arrow from a graph G to the forgetful functor U is defined as a pair $\langle C, u : G \to U(C) \rangle$, where u is a "universal" graph homomorphism. The universal arrow property asserts that every graph homomorphism $\phi : G \to H$ uniquely factors through the universal graph homomorphism $u : G \to U(C)$, where $U(C)$ is the graph induced by category C defining the universal arrow property. In other words, the associated *extension* problem of "completing" the triangle of graph homomorphisms in the category of **Graph** can be uniquely solved by "lifting" the associated category arrow $h : C \to D$.
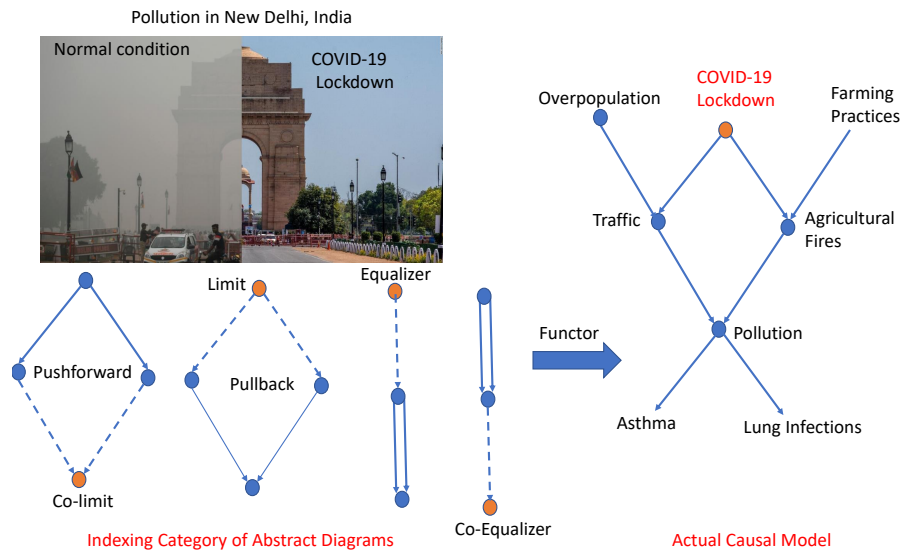

Figure 3. A causal model of climate change and COVID-19 lockdown. Universal causality defines causal models as functors mapping from an indexing category of abstract diagrams to the actual causal model.

Figure 4 illustrates the concept of causal simplicial structures. Here, X denotes a causal structure represented as a category. $X[0]$ represents the "objects" of the causal structure, defined formally as the contravariant functor $X[0] : [0] \to X$ from the simplicial category Δ to the causal category X. The arrows representing causal effects are defined as $X[1] : [1] \to X$. Note that since $[1] = \{0, 1\}$ is a category by itself, it has one (non-identity) arrow $0 \to 1$ (as well as two identity arrows). The mapping of this arrow onto X defines the "edges" of the causal model. Similarly, X_2 represents oriented "triangles" of three objects. Note that there is one edge from X_0 to X_1, labeled by s_0. This is a co-degeneracy operator from the simplicial layer that maps each object A into an identity edge $\mathbf{1}_A$. Similarly, there

are two edges marked d_0 and d_1 from X_1 to X_0. These are co-face operators that map an edge to its source and target vertices correspondingly. Notice also that there are three edges from X_2 to X_1, marked d_0, d_1, and d_2. These are the "faces" of each 2-simplex as shown. Consider the fragment of the causal DAG model from Figure 3 shown on the right in Figure 4. The *order complex* of a DAG forms a simplicial object as shown, where the simplices are represented by the nonempty chains. In particular, each path of length n defines a simplex of size n. For example, the path from O (representing **Overpopulation**) to T (representing **Traffic**) to P (representing **Pollution**) defines a simplex of size 2, shown as the green shaded triangle. Note the simplices are *oriented*, which is not shown for simplicity in Figure 4. Thus, the 2-simplex formed from the chain from **O** to **T** to **P** is oriented such that **O** "points to" **T**, which in turn "points to" **P**. This mapping from chains over DAGs to simplicial objects is a special case of a more general construction discussed later in the paper, based on constructing the *nerve* of a category that provides a faithful functor embedding any (causal) category as a simplicial object. For example, the symmetric monoidal category representations of causal DAG models [25,29,30] can be faithfully embedded as simplicial objects by constructing their nerve.
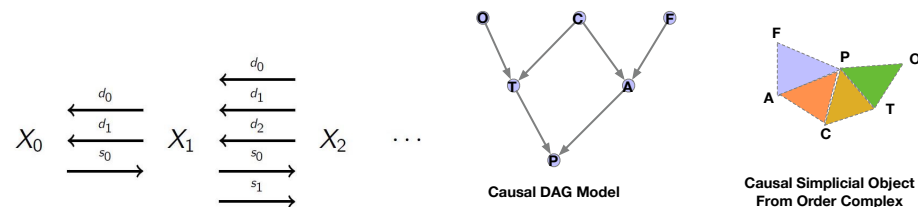

Figure 4. (**Left**) generic structure of a simplicial set. (**Right**) an oriented simplicial complex formed from the order complex of nonempty chains of the DAG model from Figure 3.

2. A Layered Architecture for Universal Causality

In this paper, we propose a layered architecture that defines the framework called UCLA (Universal Causality Layered Architecture). This architecture is illustrated in Figure 5. Table 2 describes the composition of each layer. Many variants are possible, as we will discuss in the paper. As functors compose with each other, it is also possible to consider "collapsed" versions of the UCLA hierarchy.

The UCLA architecture is built on the theoretical foundation of ordinary category theory [35–38] and higher-order category theory, including quasicategories [13], and ∞-categories [14]. As Figure 5 illustrates, at the top layer of UCLA, we model causal interventions itself as a higher-order category defined over simplicial sets and objects [12]. Causal discovery often involves a sequence of interventions, which naturally compose to form a category. Simplicial sets and simplicial objects [12] have long been a foundational framework in algebraic topology [39]. Modeling interventions using simplicial sets permits a *hierarchical* language for expressing interventions, as (co)face operators in simplicial sets and objects operate over groups of objects of arbitrary sizes. This category-theoretic approach of formalizing causal interventions gives an algebraic formalism that are related to topological notions used in causal discovery methods, such as *separating sets* [27,28] that can be defined in terms of lifting diagrams [17]. Although we will not delve into this elaboration in this paper, it is possible to define causal inference over "fuzzy" simplicial sets as well [2], which associate a real number $p \in I = (0, 1]$ with each simplicial object that denotes the uncertainty associated with a causal object or morphism. In this case, we define a fuzzy simplicial object as the functor $\Delta^{op} \times I \to \mathcal{C}$. Fuzzy simplicial sets have been recently used in data visualization [2].

Table 2. Each layer of UCLA represents a categorical abstraction of causal inference.

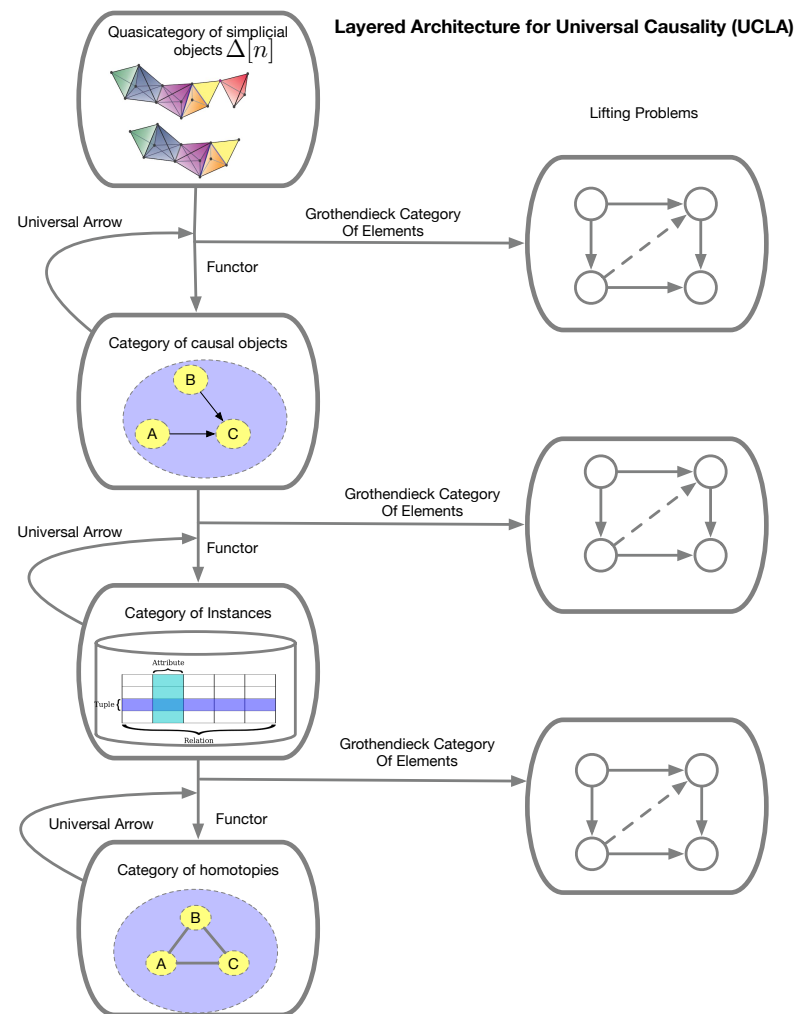| Layer	Objects	Morphisms	Description
Simplicial	$[n] = \{0, 1, \ldots, n\}$	$f = [m] \to [n]$	Category of interventions
Relational	Vertices V, Edges E	$s, t : E \to V$	Causal Model Category
Tabular	Sets	Functions on sets $f : S \to T$	Category of instances
Homotopy	Topological Spaces	Causal equivalence	Causal homotopy

Figure 5. UCLA is a layered architecture that defines Universal Causality.

The second layer of the model represents the causal category itself, which could be a causal DAG [5], a symmetric monoidal category defining a causal DAG [25], a semi-join lattice defining a conditional independence structure, such as an integer-valued multiset [8], a relational database defining a relational causal model [21], or a causal information field [6,7], which uses a measure-theoretic notion of causality. At the third layer, we model the actual data defining a causal model by a category of instances. Finally, at the bottom-most layer, we use a homotopy category to define equivalences among causal models.

The *Grothendieck Category of Elements* (GCE) is a type of universal construction [35,37,40] that plays a central role in the UCLA architecture. It is remarkably similar to other representations widely used in database theory, and specifically in the context of causal inference, it is related to the *ground graph* used in relational causal models [21,32]. However, GCE is far more general than the ground graph construction in that it can be used to embed any object or indeed any category in **Cat**, the category of all categories.

We use *lifting diagrams* [17] to formalize causal inference at each layer of the hierarchy. A lifting problem in a category $\mathcal{C}$ is a morphism $h : B \to X$ in $\mathcal{C}$ satisfying $p \circ h = \nu$ and $h \circ f = \mu$ as indicated in the commutative diagram below.

$$
\begin{array}{ccc}
A & \xrightarrow{\ \mu\ } & X \\
\downarrow{\scriptstyle f} & {\scriptstyle h}\nearrow & \downarrow{\scriptstyle p} \\
B & \xrightarrow{\ \nu\ } & Y
\end{array}
$$

Lifting diagrams were shown to be capable of expressing SQL queries in relational databases [4]. Here, we extend this approach to model causal inference under non-random interventions, exploiting the capability of the simplicial layer to impose non-random "surgical" operations on a causal category.

Finally, to explain the bottom-most layer in UCLA of homotopy categories, it is well known that causal models are not identifiable from observations alone [5]. For example, the three distinct causal DAG models over three variables $A \leftarrow B \to C$, $A \to B \to C$ and $A \leftarrow B \leftarrow C$ have the same conditional independence structure, and are equivalent given a dataset of values of the variables. To model the non-distinguishability of causal models under observation, we introduce the concept of homotopic equivalence comes from topology, and is intended to reflect equivalence under some invertible mappings. A *homotopy* from a morphism $f : X \to Y$ to another morphism $g : X \to Y$ is a continuous function $h : X \times [0, 1] \to Y$ satisfying $h(0, x) = f(x)$ and $h(x, 1) = g(x)$. In the category **Top** of topological spaces, homotopy defines an equivalence class on morphisms. In the application to causal inference, we can define causal homotopy [41] as finding the "quotient space" of the category of all causal models under a given set of invertible morphisms mapping one causal model into another equivalent model.

3. Categories, Functors, and Universal Arrows

We introduce the basic theory underlying UC in more depth now, building on relationship between categories and graphs illustrated in Figure 2. Given a graph, we can define the "free" category associated with it where we consider all possible paths between pairs of vertices (including self-loops) as the set of morphisms between them. In the reverse direction, given a category, we can define a "forgetful" functor that extracts the underlying graph from the category, forgetting the composition rule.

Definition 1. *A* **graph** $\mathcal{G}$ *(sometimes referred to as a quiver) is a labeled directed multi-graph defined by a set O of objects, a set A of arrows, along with two morphisms $s : A \to O$ and $t : A \to O$ that specify the domain and co-domain of each arrow. In this graph, we define the set of composable pairs of arrows by the set*

$$
A \times_O A = \{ \langle g, f \rangle \mid g, f \in A, \ s(g) = t(f) \}
$$

A **category** $\mathcal{C}$ *is a graph $\mathcal{G}$ with two additional functions:* $\mathbf{id} : O \to A$, *mapping each object* $c \in C$ *to an arrow* $\mathbf{id}_c$ *and* $\circ : A \times_O A \to A$, *mapping each pair of composable morphisms* $\langle f, g \rangle$ *to their composition* $g \circ f$.

It is worth emphasizing that no assumption is made here of the finiteness of a graph, either in terms of its associated objects (vertices) or arrows (edges). Indeed, it is entirely reasonable to define categories whose graphs contain an infinite number of edges. A simple example is the group $\mathbb{Z}$ of integers under addition, which can be represented as a single object, denoted $\{\bullet\}$ and an infinite number of morphisms $f : \bullet \to \bullet$, each of which represents an integer, where composition of morphisms is defined by addition. In this example, all morphisms are invertible. In a general category with more than one object, a *groupoid* defines a category all of whose morphisms are invertible.

As our paper focuses on the use of category theory to formalize causal inference, we interpret causal changes in terms of the concept of isomorphisms in category theory. We will elaborate this definition later in the paper.

Definition 2. *Two objects X and Y in a category C are deemed* **isomorphic**, *or X $\cong$ Y if and only if there is an invertible morphism f : X $\rightarrow$ Y, namely f is both* left invertible *using a morphism g : Y $\rightarrow$ X so that g $\circ$ f = id_X, and f is* right invertible *using a morphism h where f $\circ$ h = id_Y. A* **causally isomorphic change** *in a category is defined as a change of a causal object Y into $\hat{Y}$ under an intervention that changes another object X into $\hat{X}$ such that $\hat{Y} \cong Y$, that is, they are isomorphic. A* **causal non-isomorphic effect** *is a change that leads to a non-isomorphic change where $\hat{Y} \ncong Y$. An alternate definition would be to define a causally isomorphic change as a change that is an isomorphism in the category whose morphisms are causal changes.*

In the category **Sets**, two finite sets are considered isomorphic if they have the same number of elements, as it is then trivial to define an invertible pair of morphisms between them. In the category **Vect**$_k$ of vector spaces over some field *k*, two objects (vector spaces) are isomorphic if there is a set of invertible linear transformations between them. As we will see below, the passage from a set to the "free" vector space generated by elements of the set is another manifestation of the universal arrow property.

Functors can be viewed as a generalization of the notion of morphisms across algebraic structures, such as groups, vector spaces, and graphs. Functors do more than functions: they not only map objects to objects, but like graph homomorphisms, they need to also map each morphism in the domain category to a corresponding morphism in the co-domain category. Functors come in two varieties, as defined below.

Definition 3. *A* **covariant functor** *F : C $\rightarrow$ D from category C to category D, and defined as the following:*

- *An object FX (also written as F(X)) of the category D for each object X in category C.*
- *An arrow F(f) : FX $\rightarrow$ FY in category D for every arrow f : X $\rightarrow$ Y in category C.*
- *The preservation of identity and composition: F id_X = id_{FX} and (Ff)(Fg) = F(g $\circ$ f) for any composable arrows f : X $\rightarrow$ Y, g : Y $\rightarrow$ Z.*

Definition 4. *A* **contravariant functor** *F : C $\rightarrow$ D from category C to category D is defined exactly like the covariant functor, except all the arrows are reversed.*

3.1. Universal Arrows

This process of going from a category to its underlying directed graph embodies a fundamental universal construction in category theory, called the *universal arrow* [35]. It lies at the heart of many useful results, principally the Yoneda Lemma that shows how object identity itself emerges from the structure of morphisms that lead into (or out of) it. The Yoneda Lemma codifies the meaning of universal causality, as it implicitly states that any change to an object must be accompanied by a change to its presheaf structure. Consequently, we can model UC in a representation-independent manner using the Yoneda Lemma.

Definition 5. *Given a functor S : D $\rightarrow$ C between two categories, and an object c of category C, a* **universal arrow** *from c to S is a pair $\langle r, u \rangle$, where r is an object of D and u : c $\rightarrow$ Sr is an arrow of C, such that the following universal property holds true:*

- *For every pair $\langle d, f \rangle$ with d an object of D and f : c $\rightarrow$ Sd an arrow of C, there is a unique arrow f' : r $\rightarrow$ d of D with Sf' $\circ$ u = f.*

Above we used the example of functors between graphs and their associated "free" categories and graphs to illustrate universal arrows. A central principle in the UCLA architecture is that every pair of categorical layers is synchronized by a functor, along

with a universal arrow. We explore the universal arrow property more deeply in this section, showing how it provides the conceptual basis behind the Yoneda Lemma, and Grothendieck's category of elements. In the case of causal inference, universal arrows enable mimicking the effects of causal operations from one layer of the UCLA hierarchy down to the next layer. In particular, at the simplicial object layer, we can model a causal intervention in terms of face and degeneracy operators (defined below in more detail). These in turn correspond to "graph surgery" [5] operations on causal DAGs, or in terms of "copy", "delete" operators in "string diagram surgery" of causal models defined on symmetric monoidal categories [25]. These "surgery" operations at the next level may translate down to operations on probability distributions, measurable spaces, topological spaces, or chain complexes. This process follows a standard construction used widely in mathematics, for example group representations associate with any group G, a left **k**-module M representation that enables modeling abstract group operations by operations on the associated modular representation. These concrete representations must satisfy the universal arrow property for them to be faithful. A special case of the universal arrow property is that of universal element, which as we will see below plays an important role in the UCLA architecture in defining a suitably augmented category of elements, based on a construction introduced by Grothendieck.

Definition 6. *If D is a category and $H : D \to$ **Set** is a set-valued functor, a **universal element** associated with the functor H is a pair $\langle r, e \rangle$ consisting of an object $r \in D$ and an element $e \in Hr$ such that for every pair $\langle d, x \rangle$ with $x \in Hd$, there is a unique arrow $f : r \to d$ of D such that $(Hf)e = x$.*

Example 1. *Let E be an equivalence relation on a set S, and consider the quotient set S/E of equivalence classes, where $p : S \to S/E$ sends each element $s \in S$ into its corresponding equivalence class. The set of equivalence classes S/E has the property that any function $f : S \to X$ that respects the equivalence relation can be written as $fs = fs'$ whenever $s \sim_E s'$, that is, $f = f' \circ p$, where the unique function $f' : S/E \to X$. Thus, $\langle S/E, p \rangle$ is a universal element for the functor H.*

3.2. The Grothendieck Category of Elements

We turn next to define the category of elements, based on a construction by Grothendieck, and illustrate how it can serve as the basis for inference at each layer of the UCLA architecture. This definition is a special case of a general construction by Grothendieck [40].

Definition 7. *Given a set-valued functor $\delta : \mathcal{C} \to$ **Set** from some category $\mathcal{C}$, the induced* **Grothendieck category of elements** *associated with δ is a pair $(\int \delta, \pi_\delta)$, where $\int \delta \in$ **Cat** is a category in the category of all categories **Cat**, and $\pi_\delta : \int \delta \to \mathcal{C}$ is a functor that "projects" the category of elements into the corresponding original category $\mathcal{C}$. The objects and arrows of $\int \delta$ are defined as follows:*

- *$Ob(\int \delta) = \{(s, x) | x \in Ob(\mathcal{C}), x \in \delta s\}$.*
- *$\mathbf{Hom}_{\int \delta}((s, x), (s', x')) = \{f : s \to s' | \delta(f)(x) = x'\}$*

Example 2. *To illustrate the category of elements construction, let us consider the toy climate change causal model shown in Figure 6. Let the category C be defined by this causal DAG model, where the objects Ob(C) are defined by the four vertices, and the arrows $\mathbf{Hom}_C$ are defined by the four edges in the model. The set-valued functor $\delta : \mathcal{C} \to$ **Set** maps each object (vertex) in C to a set of instances, thereby turning the causal DAG model into an associated set of tables.*
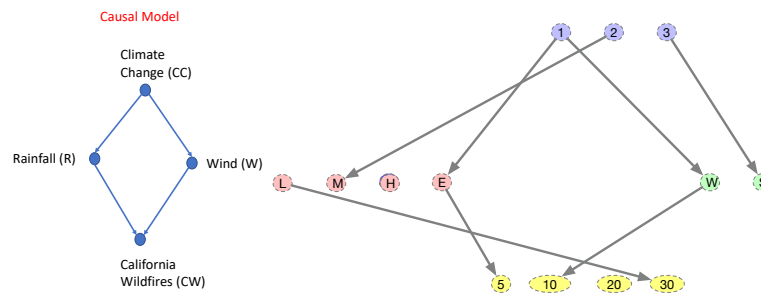
Figure 6. A toy causal DAG model of climate change to illustrate the category of elements construction. **Climate Change** is a discrete multinomial variable over three values 1, 2, and 3. For each of its values, the arrow from **Climate Change** to **Rainfall** maps each specific value of **Climate Change** to a value of **Rainfall**, thereby indicating a causal effect of climate change on the amount of rainfall in California. **Rainfall** is also a multinomial discretized as low (marked "L"), medium (marked "M"), high (marked "H"), or extreme (marked "E"). **Wind** speeds are binned into two levels (marked "W" for weak, and "S" for strong). Finally, the percentage of California wildfires is binned between 5 and 30. Not all arrows in the category of elements are shown, for clarity.

Later in the paper, we give an application of the category of elements construction to relational causal models, where in particular, it gives a rigorous semantics for ideas such as *relational skeleton* and the *ground graph* proposed in [21,32].

3.3. Yoneda Lemma

The Yoneda Lemma plays a crucial role in UC because it defines the concept of a representation in category theory. We first show that associated with universal arrows is the corresponding induced isomorphisms between **Hom** sets of morphisms in categories. This universal property then leads to the Yoneda Lemma.

Theorem 1. *Given any functor* $S : D \rightarrow C$, *the universal arrow* $\langle r, u : c \rightarrow Sr \rangle$ *implies a bijection exists between the* **Hom** *sets*
$$\mathbf{Hom}_D(r, d) \simeq \mathbf{Hom}_C(c, Sd)$$

While this is a well-known result whose proof can be found in [35], the crucial point here is its implication for causal inference. As we will see later, often in the modeling of causal inference using symmetric monoidal categories [25,29,30], a correspondence is set up between two categories, for example the symmetric monoidal category representing the structure of a causal DAG model, and the category of stochastic matrices that defines the DAG semantics. The universal arrow theorem above shows how the morphisms over the symmetric monoidal category can be synchronized with those over the stochastic matrices, enabling causal interventions to be tracked properly. A special case of this natural transformation that transforms the identity morphism $\mathbf{1}_r$ leads us to the Yoneda Lemma.

$$
\begin{array}{ccc}
D(r,r) & \xrightarrow{\phi_r} & C(c,Sr) \\
\downarrow{\scriptstyle D(r,f')} & & \downarrow{\scriptstyle C(c,Sf')} \\
D(r,d) & \xrightarrow{\phi_d} & C(c,Sd)
\end{array}
$$

As the two paths shown here must be equal in a commutative diagram, we get the property that a bijection between the **Hom** sets holds precisely when $\langle r, u : c \rightarrow Sr \rangle$ is a universal arrow from c to S. Note that for the case when the categories C and D are small, meaning their **Hom** collection of arrows forms a set, the induced functor $\mathbf{Hom}_C(c, S-)$ to **Set** is isomorphic to the functor $\mathbf{Hom}_D(r, -)$. This type of isomorphism defines a universal representation, and is at the heart of the causal reproducing property (CRP) defined below.

Lemma 1. Yoneda Lemma: *If $H : D \to$ **Set** is a set-valued functor, and r is an object in D, there is a bijection that sends each natural transformation $\alpha : \textbf{Hom}_D(r, -) \to K$ to $\alpha_r \mathbf{1}_r$, the image of the identity morphism $\mathbf{1}_r : r \to r$.*

$$y : \textbf{Nat}(\textbf{Hom}_D(r, -), K) \simeq Kr$$

The proof of the Yoneda Lemma follows directly from the below commutative diagram, a special case of the above diagram for universal arrows.

$$
\begin{array}{ccc}
D(r, r) & \xrightarrow{\phi_r} & Kr \\
\downarrow{\scriptstyle D(r, f')} & & \downarrow{\scriptstyle C(c, Sf')} \\
D(r, d) & \xrightarrow{\phi_d} & Kd
\end{array}
$$

3.4. The Universality of Diagrams and the Causal Reproducing Property

We state two key results that underly UC, and while both these results follow directly from basic theorems in category theory, their significance for causal inference is what makes them particularly noteworthy. The first result pertains to the notion of diagrams as functors, and shows that for the functor category of presheaves, which is a universal representation of causal inference, every presheaf object can be represented as a colimit of representables through the Yoneda Lemma. This result can be seen as a generalization of the very simple result in set theory that each set is a union of one element sets. The second result is the causal reproducing property, which shows that the set of all causal effects between two objects is computable from the presheaf functor objects defined by them. Both these results are abstract, and apply to any category representation of a causal model.

Diagrams play a key role in defining UC and the UCLA architecture, as has already become clear from the discussion above. We briefly want to emphasize the central role played by universal constructions involving limits and colimits of diagrams, which are viewed as functors from an indexing category of diagrams to a category. To make this somewhat abstract definition concrete, let us look at some simpler examples of universal properties, including co-products and quotients (which in set theory correspond to disjoint unions). Coproducts refer to the universal property of abstracting a group of elements into a larger one.

Before we formally the concept of limit and colimits [35], we consider some examples. These notions generalize the more familiar notions of Cartesian products and disjoint unions in the category of **Sets**, the notion of meets and joins in the category **Preord** of preorders, as well as the least upper bounds and greatest lower bounds in lattices, and many other concrete examples from mathematics.

Example 3. *If we consider a small "discrete" category $\mathcal{D}$ whose only morphisms are identity arrows, then the colimit of a functor $\mathcal{F} : \mathcal{D} \to \mathcal{C}$ is the* categorical coproduct *of $\mathcal{F}(D)$ for D, an object of category D, is denoted as*

$$Colimit_{\mathcal{D}}F = \bigsqcup_D \mathcal{F}(D)$$

*In the special case when the category C is the category **Sets**, then the colimit of this functor is simply the disjoint union of all the sets $F(D)$ that are mapped from objects $D \in \mathcal{D}$.*

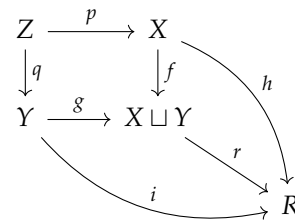**Example 4.** *Dual to the notion of colimit of a functor is the notion of* limit. *Once again, if we consider a small "discrete" category $\mathcal{D}$ whose only morphisms are identity arrows, then the limit of a functor $\mathcal{F} : \mathcal{D} \to \mathcal{C}$ is the* categorical product *of $\mathcal{F}(D)$ for D, an object of category D, is denoted as*

$$limit_{\mathcal{D}}F = \prod_D \mathcal{F}(D)$$

In the special case when the category C is the category **Sets***, then the limit of this functor is simply the Cartesian product of all the sets F(D) that are mapped from objects D ∈ 𝒟.*

Pullback and Pushforward Mappings

Universal causal models in UC are defined in terms of *universal constructions*, which satisfy a universal property. We can illustrate this concept using pullback and pushforward mappings. These notions help clarify the idea of the Grothendieck category of elements, which plays a key role in the UCLA architecture.

$$
\begin{array}{ccc}
Z & \xrightarrow{\ p\ } & X \\
\downarrow{\scriptstyle q} & & \downarrow{\scriptstyle f} \\
Y & \xrightarrow{\ g\ } & X \sqcup Y
\end{array}
$$

An example of a universal construction is given by the above commutative diagram, where the coproduct object $X \sqcup Y$ uniquely factorizes any mapping $h : X \to R$, such that any mapping $i : Y \to R$, so that $h = r \circ f$, and furthermore $i = r \circ g$. Co-products are themselves special cases of the more general notion of co-limits. Figure 7 illustrates the fundamental property of a *pullback*, which along with *pushforward*, is one of the core ideas in category theory. The pullback square with the objects U, X, Y and Z implies that the composite mappings $g \circ f'$ must equal $g' \circ f$. In this example, the morphisms f and g represent a *pullback* pair, as they share a common co-domain Z. The pair of morphisms f', g' emanating from U define a *cone*, because the pullback square "commutes" appropriately. Thus, the pullback of the pair of morphisms f, g with the common co-domain Z is the pair of morphisms f', g' with common domain U. Furthermore, to satisfy the universal property, given another pair of morphisms x, y with common domain T, there must exist another morphism $k : T \to U$ that "factorizes" x, y appropriately, so that the composite morphisms $f' k = y$ and $g' k = x$. Here, T and U are referred to as *cones*, where U is the limit of the set of all cones "above" Z. If we reverse arrow directions appropriately, we get the corresponding notion of pushforward. So, in this example, the pair of morphisms f', g' that share a common domain represent a pushforward pair. As Figure 7, for any set-valued functor $\delta : S \to$ **Sets**, the Grothendieck category of elements $\int \delta$ can be shown to be a pullback in the diagram of categories. Here, **Set**$_*$ is the category of pointed sets, and π is a projection that sends a pointed set $(X, x \in X)$ to its underlying set X.
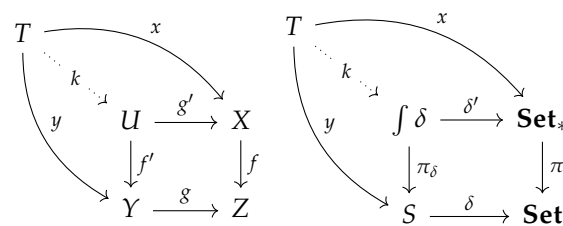
Figure 7. (**Left**) Universal Property of pullback mappings. (**Right**) The Grothendieck category of elements $\int \delta$ of any set-valued functor $\delta : S \to$ **Set** can be described as a pullback in the diagram of categories. Here, **Set**$_*$ is the category of pointed sets $(X, x \in X)$, and π is the "forgetful" functor that sends a pointed set $(X, x \in X)$ into the underlying set X.

In the category **Sets**, we know that every object (i.e., a set) X can be expressed as a coproduct of its elements $X \simeq \sqcup_{x \in X} \{x\}$, where $x \in X$. Note that we can view each element $x \in X$ as a morphism $x : \{*\} \to X$ from the one-point set to X. The categorical generalization of this result is called the *density theorem* in the theory of sheaves [36]. First, we define the key concept of a *comma category*.

Definition 8. *Let $F : \mathcal{D} \to \mathcal{C}$ be a functor from category $\mathcal{D}$ to $\mathcal{C}$. The **comma category** $F \downarrow \mathcal{C}$ is one whose objects are pairs (D, f), where $D \in \mathcal{D}$ is an object of $\mathcal{D}$ and $f \in \mathbf{Hom}_{\mathcal{C}}(F(D), C)$, where C is an object of $\mathcal{C}$. Morphisms in the comma category $F \downarrow \mathcal{C}$ from (D, f) to (D', f'), where $g : D \to D'$, such that $f' \circ F(g) = f$. We can depict this structure through the following commutative diagram:*

$$
\begin{array}{ccc}
 & F(D) & \\
 & \swarrow_{F(g)} \quad \searrow^{f} & \\
F(D') & \xrightarrow{\quad f' \quad} & C
\end{array}
$$

We first introduce the concept of a *dense* functor [40]:

Definition 9. *Let D be a small category, C be an arbitrary category, and $F : \mathcal{D} \to \mathcal{D}$ be a functor. The functor F is **dense** if for all objects C of $\mathcal{C}$, the natural transformation*

$$\psi_F^C : F \circ U \to \Delta_C, \ \ (\psi_F^C)_{(\mathcal{D}, f)} = f$$

is universal in the sense that it induces an isomorphism $Colimit_{F \downarrow C} F \circ U \simeq C$. Here, $U : F \downarrow C \to \mathcal{D}$ is the projection functor from the comma category $F \downarrow \mathcal{C}$, defined by $U(D, f) = D$.

A fundamental consequence of the category of elements is that every object in the functor category of presheaves, namely contravariant functors from a category into the category of sets, is the colimit of a diagram of representable objects, via the Yoneda Lemma. Notice this is a generalized form of the density notion from the category **Sets**, as explained above.

Theorem 2. Universality of Diagrams in UC: *In the functor category of presheaves $\mathbf{Set}^{\mathcal{C}^{op}}$, every object P is the colimit of a diagram of representable objects, in a canonical way [36].*

To explain the significance of this result for causal inference, note that UC represents causal diagrams as functors from an indexing category of diagrams to an actual causal model (as illustrated earlier in Figure 3). The density theorem above tells us that every presheaf object can be represented as a colimit of (simple) representable objects, namely functor objects of the form $\mathbf{Hom}_{\mathcal{C}}(-, c)$.

Reproducing Kernel Hilbert Spaces (RKHS's) transformed the study of machine learning, precisely because they are the unique subcategory in the category of all Hilbert spaces that have representers of evaluation defined by a kernel matrix $K(x, y)$ [42]. The reproducing property in an RKHS is defined as $\langle K(x, -), K(-, y) \rangle = K(x, y)$. An analogous but far more general reproducing property holds in the UC framework, based on the Yoneda Lemma. The significance of the Causal Reproducing Property is that presheaves act as "representers" of causal information, precisely analogous to how kernel matrices act as representers in an RKHS.

Theorem 3. Causal Reproducing Property: *All causal influences between any two objects X and Y can be derived from its presheaf functor objects, namely*

$$\mathbf{Hom}_{\mathcal{C}}(X, Y) \simeq \mathbf{Nat}(\mathbf{Hom}_{\mathcal{C}}(-, X), \mathbf{Hom}_{\mathcal{C}}(-, Y))$$

Proof. The proof of this theorem is a direct consequence of the Yoneda Lemma, which states that for every presheaf functor object F in $\hat{\mathcal{C}}$ of a category $\mathcal{C}$, $\mathbf{Nat}(\mathbf{Hom}_{\mathcal{C}}(-, X), F) \simeq FX$. That is, elements of the set FX are in $1 - 1$ bijections with natural transformations from the presheaf $\mathbf{Hom}_{\mathcal{C}}(-, X)$ to F. For the special case where the functor object $F = \mathbf{Hom}_{\mathcal{C}}(-, Y)$, we get the result immediately that $\mathbf{Hom}_{\mathcal{C}}(X, Y) \simeq \mathbf{Nat}(\mathbf{Hom}_{\mathcal{C}}(-, X), \mathbf{Hom}_{\mathcal{C}}(-, Y))$. □

In UC, any causal influence of an object X upon any other object Y can be represented as a natural transformation (a morphism) between two functor objections in the presheaf category $\hat{\mathcal{C}}$. The CRP is very akin to the idea of the reproducing property in kernel methods.

3.5. Lifting Problems

The UCLA hierarchy is defined through a series of categorical abstractions of a causal model, ranging from a combinatorial model defined by a simplicial object down to a measure-theoretic or topological realization. Between each pair of layers, we can formulate a series of lifting problems [17]. Lifting problems provide elegant ways to define basic notions in a wide variety of areas in mathematics. For example, the notion of injective and surjective functions, the notion of separation in topology, and many other basic constructs can be formulated as solutions to lifting problems. Database queries in relational databases can be defined using lifting problems [4]. Lifting problems define ways of decomposing structures into simpler pieces, and putting them back together again.

Definition 10. *Let $\mathcal{C}$ be a category. A **lifting problem** in $\mathcal{C}$ is a commutative diagram σ in $\mathcal{C}$.*

$$
\begin{array}{ccc}
A & \xrightarrow{\mu} & X \\
\downarrow{f} & & \downarrow{p} \\
B & \xrightarrow{\nu} & Y
\end{array}
$$

Definition 11. *Let $\mathcal{C}$ be a category. A **solution to a lifting problem** in $\mathcal{C}$ is a morphism $h : B \to X$ in $\mathcal{C}$ satisfying $p \circ h = \nu$ and $h \circ f = \mu$ as indicated in the diagram below.*

$$
\begin{array}{ccc}
A & \xrightarrow{\mu} & X \\
\downarrow{f} & \overset{h}{\nearrow} & \downarrow{p} \\
B & \xrightarrow{\nu} & Y
\end{array}
$$

Definition 12. *Let $\mathcal{C}$ be a category. If we are given two morphisms $f : A \to B$ and $p : X \to Y$ in $\mathcal{C}$, we say that f has the **left lifting property** with respect to p, or that p has the **right lifting property** with respect to f if for every pair of morphisms $\mu : A \to X$ and $\nu : B \to Y$ satisfying the equations $p \circ \mu = \nu \circ f$, the associated lifting problem indicated in the diagram below.*

$$
\begin{array}{ccc}
A & \xrightarrow{\mu} & X \\
\downarrow{f} & \overset{h}{\nearrow} & \downarrow{p} \\
B & \xrightarrow{\nu} & Y
\end{array}
$$

admits a solution given by the map $h : B \to X$ satisfying $p \circ h = \nu$ and $h \circ f = \mu$.

Example 5. *Given the paradigmatic non-surjective morphism $f : \varnothing \to \{\bullet\}$, any morphism p that has the right lifting property with respect to f is a **surjective mapping**. .*

$$
\begin{array}{ccc}
\varnothing & \xrightarrow{\mu} & X \\
\downarrow{f} & \overset{h}{\nearrow} & \downarrow{p} \\
\{\bullet\} & \xrightarrow{\nu} & Y
\end{array}
$$

Example 6. *Given the paradigmatic non-injective morphism $f : \{\bullet, \bullet\} \to \{\bullet\}$, any morphism p that has the right lifting property with respect to f is an **injective mapping**. .*

$$
\begin{array}{ccc}
\{\bullet, \bullet\} & \xrightarrow{\mu} & X \\
\downarrow{f} & \overset{h}{\nearrow} & \downarrow{p} \\
\{\bullet\} & \xrightarrow{\nu} & Y
\end{array}
$$

4. Universal Conditional Independence in Categories

Before proceeding to further detail the UCLA architecture, we discuss the special role played by conditional independence in causal inference. Causal models can be abstractly characterized by their underlying conditional independences. A number of previous axiomatizations such as *graphoids* [5,16], *integer-valued multisets* [8], and *separoids* [10] can be subsumed under the category-theoretic notion of universal conditional independence [15]. Conditional independence structures have been actively studied in AI, causal inference, machine learning, probability, and statistics for many years. Dawid [10] define separoids, a join semi-lattice, to formalize reasoning about conditional independence and irrelevance in many areas, including statistics. Pearl [16] introduced *graphoids*, a distributive lattice over disjoint subsets of variables, to model reasoning about irrelevance in probabilistic systems, and proposed representations using directed acyclic graphs (DAGs). Studeny [8] proposed a lattice-theoretic model of conditional independences using integer-valued multisets to address the intrinsic limitations of DAG-based representations.

In particular, we want to show how it is possible to define *universal conditional independence* [15], a representation of conditional independence in any category. We build specifically on the notion of *separoids* [10], an algebraic characterization of conditional independence. Recent work by Fritz and Klingler [30] has proposed a symmetric monoidal category representation of DAG type causal models, and an associated categorical probabilistic representations of *d*-separation. Our goals are to construct a more abstract representation of conditional independence based on non-graphical representations, like separoids [10] as well as integer-valued multisets [8].

Conditional independence plays a key role in causal discovery as it is often used as an oracle in causal discovery from data. Consider the problem of causal discovery as inferring a directed acyclic graph (DAG) $G = (V, E)$ from data, where the conditional independence $\perp\!\!\!\perp$ property is defined using the graph property of *d*-separation [16]. A given DAG G can be characterized in two ways: one parameterization specifies the DAG G in terms of the vertices V and edges E, which corresponds to specifying the objects and morphisms of a category defining the DAG. The second way to parameterize a DAG is by its induced collection of conditional independence properties, as defined by *d*-separation. For example, the serial DAG over three variables, $A \rightarrow B \rightarrow C$, can be defined using its two edges $A \rightarrow B$ and $B \rightarrow C$, but also by its conditional independences, namely $A \perp\!\!\!\perp C|B$ using the theory of *d*-separation. We are thus given two possibly redundant parameterizations of the same algebraic structure. However, multiple DAG models can define the same conditional independences. For example, the serial model $A \rightarrow B \rightarrow C$, as well as the "diverging" model $A \leftarrow B \rightarrow C$ and the "reverse" serial model $A \leftarrow B \leftarrow C$ all capture the same conditional independence property ($A \perp\!\!\!\perp C|B$). This non-uniqueness property arises because Bayes rule can be used to map any one of these three DAGs into the form represented by one of the other DAGs.

4.1. The Category of Separoids

A separoid $(\mathcal{S}, \leq, \perp\!\!\!\perp)$ [10] is defined as a semi-lattice $\mathcal{S}$, where the join $\vee$ operator over the semi-lattice $\mathcal{S}$ defines a preorder $\leq$, and the ternary relation $\perp\!\!\!\perp$ is defined over triples of the form $(x \perp\!\!\!\perp y|z)$ (which are interpreted to mean x is conditionally independent of y given z). We show briefly how to define a category for universal conditional independence, where each object is a separoid, and the morphisms are homomorphisms from one separoid to another. It is possible to define "lattice" objects in any category by interpreting an arrow $f : x \rightarrow y$ as defining the partial ordering [36].

Definition 13. *A **separoid** [10] defines a category over a preordered set $(\mathcal{S}, \leq)$, namely $\leq$ is reflexive and transitive, equipped with a* ternary *relation $\perp\!\!\!\perp$ on triples (x, y, z), where $x, y, z \in \mathcal{S}$ satisfy the following properties:*

- **S1:** *$(\mathcal{S}, \leq)$ is a join semi-lattice.*
- **P1:** *$x \perp\!\!\!\perp y \mid x$*

- **P2:** $x \perp\!\!\!\perp y \mid z \;\;\Rightarrow\;\; y \perp\!\!\!\perp x \mid z$
- **P3:** $x \perp\!\!\!\perp y \mid z \;\; and \;\; w \leq y \;\;\Rightarrow\;\; x \perp\!\!\!\perp w \mid z$
- **P4:** $x \perp\!\!\!\perp y \mid z \;\; and \;\; w \leq y \;\;\Rightarrow\;\; x \perp\!\!\!\perp y \mid (z \vee w)$
- **P5:** $x \perp\!\!\!\perp y \mid z \;\; and \;\; x \perp\!\!\!\perp w \mid (y \vee z) \;\;\Rightarrow\;\; x \perp\!\!\!\perp (y \vee w) \mid z$

*A **strong separoid** also defines a categoroid. A strong separoid is defined over a lattice $\mathcal{S}$ has in addition to a join $\vee$, a meet $\wedge$ operation, and satisfies an additional axiom:*

- **P6**: *If $z \leq y$ and $w \leq y$, then $x \perp\!\!\!\perp y \mid z \;\; and \;\; x \perp\!\!\!\perp y \mid w \;\;\Rightarrow\;\; x \perp\!\!\!\perp y \mid z \wedge w$*

To define a category of separoids, we have to define the notion of a homomorphism between separoids [10]:

Definition 14. *Let $\langle \mathcal{S}, \leq, \perp\!\!\!\perp \rangle$ and $\langle \mathcal{S}', \leq', \perp\!\!\!\perp' \rangle$ be two separoids. A map $f : \mathcal{S} \rightarrow \mathcal{S}'$ is a **separoid homomorphism** if:*

1. *It is a join-lattice homomorphism, namely $f(x \vee y) = f(x) \vee' f(y)$, which implies that $x \leq y \rightarrow f(x) \leq' f(y)$.*
2. *$x \perp\!\!\!\perp y \mid z \rightarrow f(x) \perp\!\!\!\perp' f(y) \mid f(z)$.*
3. *In case both S and S' are strong separoids, we can define the notion of a strong separoid homomorphism to additionally include the condition: $f(x \wedge y) \rightarrow f(x) \wedge' f(y)$.*

With this definition, we can now define the category of separoids and a representation-independent characterization of universal conditional independence as follows:

Theorem 4. *The category of separoids is defined as one where each object in the category is defined as a separoid $\langle \mathcal{S}, \leq, \perp\!\!\!\perp \rangle$, and the arrows are defined as (strong) separoid homomorphisms. The category of separoids provides an axiomatization of universal conditional independence, namely that it enables a universal representation through the use of universal arrows and Yoneda Lemma.*

Proof. First, we note that the category of separoids indeed forms a category as it straightforwardly satisfies all the basic properties. The (strong) separoid homomorphisms compose, so that $g \circ f$ as a composition of two (strong) separoid homomorphisms produces another (strong) separoid homomorphism. The universal property derives from the use of the Yoneda Lemma to define a category of presheaves that map from the category of separoids to the category **Sets**. $\square$

4.2. Adjoint Functors in Causal Discovery

First, we need to review the basic concept of adjoint functors, which will be helpful in modeling several aspects of causal inference in this paper.

Definition 15. *A pair of **adjoint functors** is defined as $F : \mathcal{C} \rightarrow \mathcal{D}$ and $G : \mathcal{D} \rightarrow \mathcal{C}$, where F is considered the right adjoint, and G is considered the left adjoint,*

$$\mathcal{D} \; \underset{F}{\overset{G}{\underset{\top}{\rightleftarrows}}} \; \mathbf{C}.$$

must satisfy the property that for each pair of objects C of $\mathcal{C}$ and D of $\mathcal{D}$, there is a natural transformation between the two sets of morphisms

$$\phi_{C,D} : \mathbf{Hom}_{\mathcal{C}}(C, G(D)) \simeq \mathbf{Hom}_{\mathcal{D}}(F(C), D)$$

An important property of adjoint functors is connected to the concepts of limits and colimits reviewed above.

Theorem 5. *If F and G are a pair of adjoint functors*

$$\mathcal{D} \xrightarrow[\substack{\leftarrow \\ F}]{\substack{G \\ \top}} \mathbf{C}.$$

then the functor G preserves colimits and the functor F preserves limits.

Notice the similarity of this definition to the one earlier where the universal arrow property induced a bijection of **Hom** sets that then led to universal elements, Grothendieck category of elements, and the Yoneda Lemma.

We now introduce the perspective of adjoint functors for causal discovery (see Figure 8). Many causal discovery algorithms [27] that use a conditional independence oracle to query conditional independence properties from a dataset can be viewed in this perspective as using adjoint functors between the category of separoids and the category of the causal model itself. We can design functors that map from the category of all separoids into the category of causal models (in particular, for example, the category of graphs, or the category of integer-valued multisets [8]). Shown in the figure is one particular separoid object with a single conditional independence property stating that A and B are dependent conditional on knowing the value of C), which can realized in two ways: one using a collider DAG $A \rightarrow C \leftarrow B$, and the other as a integer-valued multiset. These pair of functors are an example of the general case of adjoint functors between "forgetful" and "free" functors [40]. To make this more precise, let us define the "forgetful" functor R between a causal model on the right to its underlying set of conditional independences on the left, so that $R(M)$ is the separoid object that represents the conditional independence in a causal model M. Note that R is a "forgetful" functor, in that it "throws away" structural information, including for example, whether the causal model is a causal DAG or an integer-valued multiset. On the other hand, the "free" functor $L(M)$, its left adjoint, maps a given separoid object to any of its associated "free" objects, namely causal models that represent it, irrespective of their formalism. Within the category of causal models, morphisms enable translation between different representations.
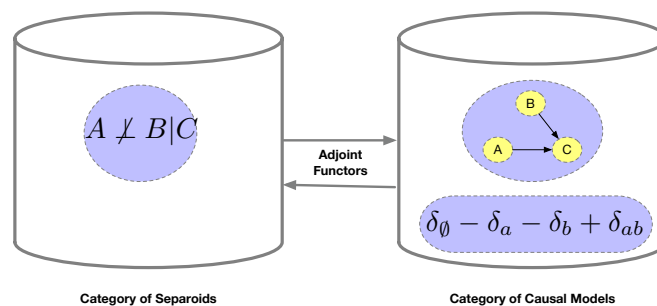

Figure 8. Adjoint functors between the category of separoids and the category of causal models. Here, a causal "collider" DAG over three random variables A, B, and C, and its associated integer-valued multiset, can both be viewed as "free" objects associated with a separoid conditional independence object, whereas the latter can be viewed in terms of a forgetful functor that throws away the causal DAG or integer-valued multiset structure.

5. Layers 1 and 2: Category of Causal Interventions over Simplicial Objects

We now discuss Layers 1 and 2 in UCLA architecture, describing the top simplicial objects layer, and how it interacts with the causal category structure (layer 2). Simplicial sets are higher-dimensional generalizations of directed graphs, partially ordered sets, as well as regular categories themselves. Importantly, simplicial sets and simplicial objects form a foundation for higher-order category theory [13,14]. By using simplicial sets and objects at the top layer, UCLA enables a powerful machinery to define a higher-order category for representing a rich class of causal interventions over a very expressive set of causal

models, including relational causal models [32], and perform abstract "diagram surgery", for example "graph surgery" [5] or "string diagram surgery" [25].

Simplicial objects have long been a foundation for algebraic topology [12,39], and more recently in higher-order category theory [13,14,43]. The category Δ has non-empty ordinals $[n] = \{0, 1, \ldots, n\}$ as objects, and order-preserving maps $[m] \to [n]$ as arrows. An important property in Δ is that any many-to-many mapping is decomposable as a composition of an injective and a surjective mapping, each of which is decomposable into a sequence of elementary injections $\delta_i : [n] \to [n+1]$, called *coface* mappings, which omits $i \in [n]$, and a sequence of elementary surjections $\sigma_i : [n] \to [n-1]$, called *co-degeneracy* mappings, which repeats $i \in [n]$. The fundamental simplex $\Delta([n])$ is the presheaf of all morphisms into $[n]$, that is, the representable functor $\Delta(-, [n])$. The Yoneda Lemma [35] assures us that an n-simplex $x \in X_n$ can be identified with the corresponding map $\Delta[n] \to X$. Every morphism $f : [n] \to [m]$ in Δ is functorially mapped to the map $\Delta[m] \to \Delta[n]$ in $\mathcal{S}$.

Any morphism in the category Δ can be defined as a sequence of *co-degeneracy* and *co-face* operators, where the co-face operator $\delta_i : [n-1] \to [n], 0 \le i \le n$ is defined as:

$$\delta_i(j) = \begin{cases} j, & \text{for } 0 \le j \le i-1 \\ j+1 & \text{for } i \le j \le n-1 \end{cases}$$

Analogously, the co-degeneracy operator $\sigma_j : [n+1] \to [n]$ is defined as

$$\sigma_j(k) = \begin{cases} j, & \text{for } 0 \le k \le j \\ k-1 & \text{for } j < k \le n+1 \end{cases}$$

Note that under the contravariant mappings, co-face mappings turn into face mappings, and co-degeneracy mappings turn into degeneracy mappings. That is, for any simplicial object (or set) X_n, we have $X(\delta_i) := d_i : X_n \to X_{n-1}$, and likewise, $X(\sigma_j) := s_j : X_{n-1} \to X_n$.

The compositions of these arrows define certain well-known properties [12,40]:

$$\begin{aligned} \delta_j \circ \delta_i &= \delta_i \circ \delta_{j-1}, \ i < j \\ \sigma_j \circ \sigma_i &= \sigma_i \circ \sigma_{j+1}, \ i \le j \end{aligned}$$

$$\sigma_j \circ \delta_i(j) = \begin{cases} \sigma_i \circ \sigma_{j+1}, & \text{for } i < j \\ 1_{[n]} & \text{for } i = j, j+1 \\ \sigma_{i-1} \circ \sigma_j, \text{for } i > j+1 \end{cases}$$

Example 7. *The "vertices" of a simplicial object C_n are the objects in C, and the "edges" of C are its arrows $f : X \to Y$, where X and Y are objects in C. Given any such arrow, the degeneracy operators $d_0 f = Y$ and $d_1 f = X$ recover the source and target of each arrow. Also, given an object X of category C, we can regard the face operator $s_0 X$ as its identity morphism $\mathbf{1}_X : X \to X$.*

Example 8. *Given a category C, we can identify an n-simplex σ of a simplicial set C_n with the sequence:*

$$\sigma = C_o \xrightarrow{f_1} C_1 \xrightarrow{f_2} \ldots \xrightarrow{f_n} C_n$$

the face operator d_0 applied to σ yields the sequence

$$d_0 \sigma = C_1 \xrightarrow{f_2} C_2 \xrightarrow{f_3} \ldots \xrightarrow{f_n} C_n$$

where the object C_0 is "deleted" along with the morphism f_0 leaving it. The "edge intervention" model in [44] effectively can be viewed as deleting the vertex from which the edge originates.

Example 9. *Given a category $\mathcal{C}$, and an n-simplex σ of the simplicial set $\mathcal{C}_n$, the face operator d_n applied to σ yields the sequence*

$$d_n\sigma = C_0 \xrightarrow{f_1} C_1 \xrightarrow{f_2} \ldots \xrightarrow{f_{n-1}} C_{n-1}$$

where the object C_n is "deleted" along with the morphism f_n entering it. Note this face operator can be viewed as analogous to interventions on leaf nodes in a causal DAG model.

Example 10. *Given a category $\mathcal{C}$, and an n-simplex σ of the simplicial set $\mathcal{C}_n$ the face operator $d_i, 0 < i < n$ applied to σ yields the sequence*

$$d_i\sigma = C_0 \xrightarrow{f_1} C_1 \xrightarrow{f_2} \ldots C_{i-1} \xrightarrow{f_{i+1}\circ f_i} C_{i+1} \ldots \xrightarrow{f_n} C_n$$

where the object C_i is "deleted" and the morphisms f_i is composed with morphism f_{i+1}. Note that this process can be abstractly viewed as intervening on object C_i by choosing a specific value for it (which essentially "freezes" the morphism f_i entering object C_i to a constant value).

Example 11. *Given a category $\mathcal{C}$, and an n-simplex σ of the simplicial set $\mathcal{C}_n$, the degeneracy operator $s_i, 0 \leq i \leq n$ applied to σ yields the sequence*

$$s_i\sigma = C_0 \xrightarrow{f_1} C_1 \xrightarrow{f_2} \ldots C_i \xrightarrow{\mathbf{1}_{C_i}} C_i \xrightarrow{f_{i+1}} C_{i+1} \ldots \xrightarrow{f_n} C_n$$

where the object C_i is "repeated" by inserting its identity morphism $\mathbf{1}_{C_i}$.

Definition 16. *Given a category $\mathcal{C}$, and an n-simplex σ of the simplicial set $\mathcal{C}_n$, σ is a **degenerate** simplex if some f_i in σ is an identity morphism, in which case C_i and C_{i+1} are equal.*

5.1. Simplicial Subsets and Horns

We now describe more complex ways of extracting parts of causal structures using simplicial subsets and horns. These structures will play a key role in defining suitable lifting problems.

Definition 17. *The **standard simplex** Δ^n is the simplicial set defined by the construction*

$$([m] \in \Delta) \mapsto \mathbf{Hom}_\Delta([m], [n])$$

By convention, $\Delta^{-1} := \varnothing$. The standard 0-simplex Δ^0 maps each $[n] \in \Delta^{op}$ to the single element set $\{\bullet\}$.

Definition 18. *Let $S_\bullet$ denote a simplicial set. If for every integer $n \geq 0$, we are given a subset $T_n \subseteq S_n$, such that the face and degeneracy maps*

$$d_i : S_n \to S_{n-1} \quad s_i : S_n \to S_{n+1}$$

applied to T_n result in

$$d_i : T_n \to T_{n-1} \quad s_i : T_n \to T_{n+1}$$

*then the collection $\{T_n\}_{n \geq 0}$ defines a **simplicial subset** $T_\bullet \subseteq S_\bullet$.*

Definition 19. *The **boundary** is a simplicial set $(\partial\Delta^n) : \Delta^{op} \to \mathbf{Set}$ defined as*

$$(\partial\Delta^n)([m]) = \{\alpha \in \mathbf{Hom}_\Delta([m], [n]) : \alpha \text{ is not surjective}\}$$

Note that the boundary $\partial\Delta^n$ is a simplicial subset of the standard n-simplex Δ^n.

Definition 20. *The* **Horn** $\Lambda_i^n : \Delta^{op} \to$ **Set** *is defined as*

$$(\Lambda_i^n)([m]) = \{\alpha \in \mathbf{Hom}_\Delta([m],[n]) : [n] \nsubseteq \alpha([m]) \cup \{i\}\}$$

Intuitively, the Horn Λ_i^n can be viewed as the simplicial subset that results from removing the interior of the n-simplex Δ^n together with the face opposite its ith vertex.

5.2. Example: Causal Intervention and Horn Filling of Simplicial Objects

Let us illustrate this abstract discussion above by instantiating it in the context of causal inference. Figure 9 instantiates the abstract discussion above in terms of an example from causal inference. We are given a simple 3 variable DAG, on which we desire to explore the causal effect of variable A on C. Using Pearl's backdoor criterion, we can intervene on variable A by freezing its value $do(A = 1)$, for example, which will eliminate the dependence of A on B. Consider now the lifting problem where we want to know if there is a completion of this simplicial subset Λ_2^2, which is a "outer horn".

Figure 9. Causal interventions can be related to horns of a simplicial object.

We can view the causal intervention problem in the more abstract setting of a class of lifting problem, shown with the following diagrams. Consider the problem of composing 1-dimensional simplices to form a 2-dimensional simplicial object. Each simplicial subset of an n-simplex induces a a *horn* Λ_k^n, where $0 \le k \le n$. Intuitively, a horn is a subset of a simplicial object that results from removing the interior of the n-simplex and the face opposite the ith vertex. Consider the three horns defined below. The dashed arrow --→ indicates edges of the 2-simplex Δ^2 not contained in the horns.
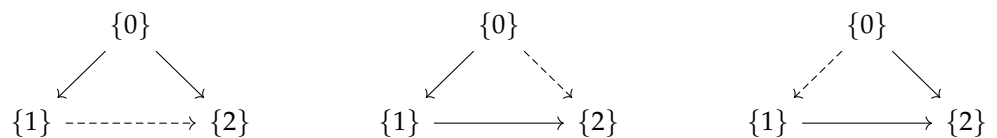
The inner horn Λ_1^2 is the middle diagram above, and admits an easy solution to the "horn filling" problem of composing the simplicial subsets. The two outer horns on either end pose a more difficult challenge. For example, filling the outer horn Λ_0^2 when the morphism between $\{0\}$ and $\{1\}$ is f and that between $\{0\}$ and $\{2\}$ is the identity $\mathbf{1}$ is tantamount to finding the left inverse of f up to homotopy. Dually, in this case, filling the outer horn Λ_2^2 is tantamount to finding the right inverse of f up to homotopy. A considerable elaboration of the theoretical machinery in category theory is required to describe the various solutions proposed, which led to different ways of defining higher-order category theory [13,14,43].

5.3. Higher-Order Categories

We now formally introduce higher-order categories, building on the framework proposed in a number of formalisms [13,14,43]. We briefly summarize various approaches to the horn filling problem in higher-order category theory.

Definition 21. *Let $f : X \to S$ be a morphism of simplicial sets. We say f is a* **Kan fibration** *if, for each $n > 0$, and each $0 \le i \le n$, every lifting problem.*

$$\begin{array}{ccc} \Lambda_i^n & \xrightarrow{\sigma_0} & X \\ \downarrow & \overset{\sigma}{\nearrow} & \downarrow f \\ \Delta^n & \xrightarrow{\bar{\sigma}} & S \end{array}$$

admits a solution. More precisely, for every map of simplicial sets $\sigma_0 : \Lambda_i^n \to X$ and every n-simplex $\bar{\sigma} : \Delta^n \to S$ extending $f \circ \sigma_0$, we can extend σ_0 to an n-simplex $\sigma : \Delta^n \to X$ satisfying $f \circ \sigma = \bar{\sigma}$.

Example 12. *Given a simplicial set X, then a projection map $X \to \Delta^0$ that is a Kan fibration is called a* **Kan complex**.

Example 13. *Any isomorphism between simplicial sets is a Kan fibration.*

Example 14. *The collection of Kan fibrations is closed under retracts.*

Definition 22 ([14]). *An ∞-category is a simplicial object $S_\bullet$ which satisfies the following condition:*

- *For $0 < i < n$, every map of simplicial sets $\sigma_0 : \Lambda_i^n \to S_\bullet$ can be extended to a map $\sigma : \Delta^n \to S_i$.*

This definition emerges out of a common generalization of two other conditions on a simplicial set S_i:

1. **Property K**: For $n > 0$ and $0 \leq i \leq n$, every map of simplicial sets $\sigma_0 : \Lambda_i^n \to S_\bullet$ can be extended to a map $\sigma : \Delta^n \to S_i$.
2. **Property C**: for $0 < 1 < n$, every map of simplicial sets $\sigma_0 : \Lambda_i^n \to S_i$ can be extended uniquely to a map $\sigma : \Delta^n \to S_i$.

Simplicial objects that satisfy property K were defined above to be Kan complexes. Simplicial objects that satisfy property C above can be identified with the nerve of a category, which yields a full and faithful embedding of a category in the category of sets. Definition 22 generalizes both of these definitions, and was called a *quasicategory* in [13] and *weak Kan complexes* in [43] when $\mathcal{C}$ is a category. We will use the nerve of a category below in defining homotopy colimits as a way of characterizing a causal model.

5.4. Example: Simplicial Objects over Integer-Valued Multisets

To help ground out this somewhat abstract discussion above on simplicial objects and sets, let us consider its application to two other examples. Our first example comes from a non-graphical representations of conditional independence, namely integer-valued multisets [8], defined as an integer-valued multiset function $u : \mathbb{Z}^{\mathcal{P}(\mathbb{Z})} \to \mathbb{Z}$ from the power set of integers, $\mathcal{P}(\mathbb{Z})$ to integers $\mathbb{Z}$. An imset is defined over partialy ordered set (poset), defined as a distributive lattice of disjoint (or non-disjoint) subsets of variables. The bottom element is denoted $\varnothing$, and top element represents the complete set of variables N. A full discussion of the probabilistic representations induced by imsets is given [8]. We will only focus on the aspects of imsets that relate to its conditional independence structure, and its topological structure as defined by the poset. A *combinatorial* imset is defined as:

$$u = \sum_{A \subset N} c_A \delta_A$$

where c_A is an integer, δ_A is the characteristic function for subset A, and A potentially ranges over all subsets of N. An *elementary* imset is defined over $(a, b \perp\!\!\!\perp A)$, where a, b are singletons, and $A \subset N \setminus \{a, b\}$. A *structural* imset is defined as one where the coefficients can be rational numbers. For a general DAG model $G = (V, E)$, an imset in standard form [8] is defined as

$$u_G = \delta_V - \delta_\varnothing + \sum_{i \in V} (\delta_{\mathbf{Pa}_i} - \delta_{i \cup \mathbf{Pa}_i})$$

The space of all possible imset representations over *n* variables defines a lattice [8], where the top of the lattice corresponds to the "discrete" causal model with no non-trivial morphisms, and the bottom of the lattice corresponds to the complete model with morphisms between every pair of objects. Each candidate imset defines a *causal horn*, a simplicial subobject of the complete simplex, and the process of causal structure discovery can be viewed in terms of the abstract horn filling problem defined above for higher-order categories.

5.5. Example: Simplicial Objects over String Diagrams

We now illustrate the above formalism of simplicial objects by illustrating how it applies to the special case where causal models are defined over symmetric monoidal categories [25,29,30]. For a detailed overview of symmetric monoidal categories, we recommend the book-length treatment by Fong and Spivak [1]. Symmetric monoidal categories (SMCs) are useful in modeling processes where objects can be combined together to give rise to new objects, or where objects disappear. For example, Coecke et al. [45] propose a mathematical framework for resources based on SMCs. We focus on the work of Jacobs et al. [25]. It is important to point out that monoidal categories can be defined as a special type of Grothendieck fibration [40]. We discuss one specific case of the general Grothendieck construction in the next section construction, and refer the reader to [40] for how the structure of monoidal categories itself emerges from this construction.

Our goal in this section is to illustrate how we can define simplicial objects over the SMC category CDU category **Syn**$_G$ constructed by Jacobs et al. [25] to mimic the process of working with an actual Bayesian network DAG *G* For the purposes of our illustration, it is not important to discuss the intricacies involved in this model, for which we refer the reader to the original paper. Our goal is to show that by encapsulating their SMC category in the UCLA framework, we can extend their approach as described below. In particular, we can solve an associated lifting problem that is defined by the functor mapping the simplicial category Δ to their SMC category. They use the category of stochastic matrices to capture the process of working with the joint distribution as shown in the figure. Instead, we show that one can use some other category, such as the category of **Sets**, or **Top** (the category of topological spaces), or indeed, the category **Meas** of measurable spaces.

Recall that Bayesian networks [16] define a joint probability distribution

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{Pa}(X_i)),$$

where $\mathrm{Pa}(X_i) \subset \{X_1, \ldots, X_n\} \setminus X_i$ represents a subset of variables (not including the variable itself). Jacobs et al. [25] show Bayesian network models can be constructed using symmetric monoidal categories, where the tensor product operation is used to combine multiple variables into a "tensored" variable that then probabilistically maps into an output variable. In particular, the monoidal category **Stoch** has as objects finite sets, and morphisms $f : A \to B$ are $|B| \times |A|$ dimensional stochastic matrices. Composition of stochastic matrices corresponds to matrix multiplication. The monoidal product $\otimes$ in **Stoch** is the cartesian product of objects, and the Kronecker product of matrices $f \otimes g$. Jacobs et al. [25] define three additional operations, the copy map, the discarding map, and the uniform state.

Definition 23. *A CDU category (for copy, discard, and uniform) is a SMC category* (**C**, $\otimes$, *I*), *where each object A has a copy map* $C_A : A \to A \otimes A$, *and discarding map* $D_A : A \to I$, *and a uniform state map* $U_A : I \to A$, *satisfying a set of equations detailed in Jacobs et al. [25]. CDU functors are symmetric monoidal functors between CDU categories, preserving the CDU maps.*

The key theorem we are interested in is the following from the original paper [25]:

Theorem 6. *There is an isomorphism (1-1 correspondence) between Bayesian networks based on a DAG G and CDU functors* $F : \mathbf{Syn}_G \to \mathbf{Stoch}$.

The significance of this theorem for the UCLA architecture is that it shows how the SMC category of CDU objects can be defined as Layer 2 of the UCLA hierarchy, whereas the category **Stoch** can be viewed as instantiating the Layer 3 of the UCLA hierarchy. Notice that this theorem in effect defines a universal arrow between the CDU category and the category of stochastic matrices, which is a central unifying principle in UC.

5.6. Nerve of a Category

An important concept that will play a key role in Layer 4 of the UCLA hierarchy is that of the *nerve* of a category [40]. The nerve of a category $\mathcal{C}$ enables embedding $\mathcal{C}$ into the category of simplicial objects, which is a fully faithful embedding.

Definition 24. *Let* $\mathcal{F} : \mathcal{C} \to \mathcal{D}$ *be a functor from category C to category D. If for all arrows f the mapping* $f \to Ff$

- *injective, then the functor* $\mathcal{F}$ *is defined to be* **faithful***.*
- *surjective, then the functor* $\mathcal{F}$ *is defined to be* **full***.*
- *bijective, then the functor* $\mathcal{F}$ *is defined to be* **fully faithful***.*

Definition 25. *The* **nerve** *of a category* $\mathcal{C}$ *is the set of composable morphisms of length n, for* $n \geq 1$*. Let* $N_n(\mathcal{C})$ *denote the set of sequences of composable morphisms of length n.*

$$\{C_o \xrightarrow{f_1} C_1 \xrightarrow{f_2} \dots \xrightarrow{f_n} C_n \mid C_i \text{ is an object in } \mathcal{C}, f_i \text{ is a morphism in } \mathcal{C}\}$$

The set of *n*-tuples of composable arrows in C, denoted by $N_n(\mathcal{C})$, can be viewed as a functor from the simplicial object $[n]$ to $\mathcal{C}$. Note that any nondecreasing map $\alpha : [m] \to [n]$ determines a map of sets $N_m(\mathcal{C}) \to N_n(\mathcal{C})$. The nerve of a category C is the simplicial set $N_\bullet : \Delta \to N_n(\mathcal{C})$, which maps the ordinal number object $[n]$ to the set $N_n(\mathcal{C})$.

The importance of the nerve of a category comes from a key result [40], showing it defines a full and faithful embedding of a category:

Theorem 7. *The* **nerve functor** $N_\bullet : \mathbf{Cat} \to \mathbf{Set}$ *is fully faithful. More specifically, there is a bijection* θ *defined as:*

$$\theta : \mathbf{Cat}(\mathcal{C}, \mathcal{C}') \to \mathbf{Set}_\Delta(N_\bullet(\mathcal{C}), N_\bullet(\mathcal{C}'))$$

Using this concept of a nerve of a category, we can now state a theorem that shows it is possible to easily embed the CDU symmetric monoidal category defined above that represents Bayesian Networks and their associated "string diagram surgery" operations for causal inference as a simplicial set.

Theorem 8. *Define the* **nerve** *of the CDU symmetric monoidal category* $(\mathbf{C}, \otimes, I)$*, where each object A has a copy map* $C_A : A \to A \otimes A$*, and discarding map* $D_A : A \to I$*, and a uniform state map* $U_A : I \to A$ *as the set of composable morphisms of length n, for* $n \geq 1$*. Let* $N_n(\mathcal{C})$ *denote the set of sequences of composable morphisms of length n.*

$$\{C_o \xrightarrow{f_1} C_1 \xrightarrow{f_2} \dots \xrightarrow{f_n} C_n \mid C_i \text{ is an object in } \mathcal{C}, f_i \text{ is a morphism in } \mathcal{C}\}$$

The associated **nerve functor** $N_\bullet : \mathbf{Cat} \to \mathbf{Set}$ *from the CDU category is fully faithful. More specifically, there is a bijection* θ *defined as:*

$$\theta : \mathbf{Cat}(\mathcal{C}, \mathcal{C}') \to \mathbf{Set}_\Delta(N_\bullet(\mathcal{C}), N_\bullet(\mathcal{C}'))$$

This theorem is just a special case of the above theorem attesting to the full and faithful embedding of any category using its nerve, which then makes it a simplicial set. We can then use the theoretical machinery at the top layer of the UCLA architecture to manipulate causal interventions in this category using face and degeneracy operators as defined above.

Note that the functor G from a simplicial object X to a category $\mathcal{C}$ can be lossy. For example, we can define the objects of $\mathcal{C}$ to be the elements of X_0, and the morphisms of $\mathcal{C}$ as the elements $f \in X_1$, where $f : a \to b$, and $d_0 f = a$, and $d_1 f = b$, and $s_0 a, a \in X$ as defining the identity morphisms $\mathbf{1}_a$. Composition in this case can be defined as the free algebra defined over elements of X_1, subject to the constraints given by elements of X_2. For example, if $x \in X_2$, we can impose the requirement that $d_1 x = d_0 x \circ d_2 x$. Such a definition of the left adjoint would be quite lossy because it only preserves the structure of the simplicial object X up to the 2-simplices. The right adjoint from a category to its associated simplicial object, in contrast, constructs a full and faithful embedding of a category into a simplicial set. In particular, the nerve of a category is such a right adjoint.

6. Layers 2 and 3 of UCLA: The Category of Elements in Causal Inference

Next, we turn to describe the second (from top) and third layers of the UCLA architecture, which pertain to the category of causal models (for example, a graph or a symmetric monoidal category), and the database of instances that support causal inferences. Drawing on the close correspondences between between categories and relational database schemes (see [4] for details), we can view causal queries over data as analogous to database queries, which can then be formulated by corresponding lifting problems. That is, each object in the model, e.g., a variable indicating a patient, maps into actual patients, and a variable indicating outcomes from COVID-19 exposure, maps into actual outcomes for that individual. The causal arrow from the patient variable into the exposure variable then maps into actual arrows for each patient. Causal queries of exposure to COVID-19 then become similar to database queries. In the next section, we will generalize this perspective, showing that we can map into a topological category and answer more abstract questions relating to the geometry of a dataset, or map into a category of measurable spaces to answer probabilistic queries. The structure of the lifting problem remains the same, what changes are the specifics of the underlying categories.

6.1. Grothendieck Category of Elements in Relational Causal Models

The Grothendieck category of elements is related to the notion of ground graphs used in relational causal models [32]. Using the example in their papers, we are given three generic objects, **Employee**, **Product**, and **Business-Unit**, and several morphisms, including **Develops**, **Funds**, **Salary**, **Competence**, **Revenue** and **Budget**. We can view a relational schema as shown as a category, following the approach shown in [46,47]. Note each object, such as **Employee**, maps using a functor into the category **Sets** into actual employees, such as **Paul** or **Sally**. Each morphism in the category, for example **Develops** must accordingly also be mapped by this functor into a set-valued function. So, as illustrated, we have that **Sally** is involved in developing a **Laptop**, and **Paul** is involved in developing a **Case**, both of which of course are instances of **Product**. The GCE for this relational causal model is strongly related to the so-called *relational skeleton* and *ground graph* explored in relational causal models [21,32].

A full discussion of these connections is beyond the scope of this paper, but there are some interesting differences to be noted. In their approach, relations such as **Develops** are depicted as undirected, whereas in our case, we model these as directional properties (which seems natural in this example). Ahsan et al. [32] develop a notion of *relational d-separation* in their work, and it would be interesting to construct a categorified version of that notion, an interesting problem for future work. We turn instead to discuss how GCE plays a key role in lifting problems associated with causal inference in UCLA. These provide a rigorous semantics to their use in relational causal models as well, which might be a fruitful avenue to explore in subsequent work.

6.2. Lifting Problems in Causal Inference

Many properties of Grothendieck's construction can be exploited (some of these are discussed in the context of relational database queries in [4]), but for our application to causal inference, we are primarily interested in the associated class of lifting problems that define queries in a causal model.

Definition 26. *If S is a collection of morphisms in category $\mathcal{C}$, a morphism $f : A \to B$ has the* **left lifting property with respect to S** *if it has the left lifting property with respect to every morphism in S. Analogously, we say a morphism $p : X \to Y$ has the* **right lifting property with respect to S** *if it has the right lifting property with respect to every morphism in S.*

We now turn to sketch some examples of the application of lifting problems for causal inference. Many problems in causal inference on graphs involve some particular graph property. To formulate it as a lifting problem, we will use the following generic template, following the initial application of lifting problems to database queries proposed by Spivak [4].

$$
\begin{array}{ccc}
Q & \xrightarrow{\ \mu\ } & \int \delta \\
{\scriptstyle f}\downarrow & {\scriptstyle h}\nearrow & \downarrow{\scriptstyle p} \\
R & \xrightarrow{\ \nu\ } & \mathcal{C}
\end{array}
$$

Here, Q is a generic query that we want answered, which could range from a database query, as in the original setting studied by Spivak [4], but more interestingly, it could be a particular graph property relating to causal inference (as illustrated by the following two examples), but as we will show later, it could also be related to the combinatorial category of simplicial objects used to model causal intervention, and finally, it could also be related to questions relating to the evaluation of causal models using a measure-theoretic or probability space. By suitably modifying the base category, the lifting problem formulation can be used to encode a diverse variety of problems in causal inference. R represents a fragment of the complete causal model $\mathcal{C}$, and δ is the category of elements defined above. Finally, h gives all solutions to the lifting problem. Some examples will help clarify this concept.

Example 15. *Consider the category of directed graphs defined by the category $\mathcal{G}$, where $Ob(\mathcal{G}) = \{V, E\}$, and the morphisms of $\mathcal{G}$ are given as $\mathbf{Hom}_{\mathcal{G}} = \{s, t\}$, where $s : E \to V$ and $t : E \to V$ define the source and terminal nodes of each vertex. Then, the category of all directed graphs is precisely defined by the category of all functors $\delta : \mathcal{G} \to \mathbf{Set}$. Any particular graph is defined by the functor $X : \mathcal{G} \to \mathbf{Set}$, where the function $X(s) : X(E) \to X(V)$ assigns to every edge its source vertex. For causal inference, we may want to check some property of a graph, such as the property that every vertex in X is the source of some edge. The following lifting problem ensures that every vertex has a source edge in the graph. The category of elements $\int \delta$ shown below refers to a construction introduced by Grothendieck, which will be defined in more detail later.*

$$
\begin{array}{ccc}
V(\bullet) & \xrightarrow{\ \mu\ } & \int \delta \\
{\scriptstyle f}\downarrow & {\scriptstyle h}\nearrow & \downarrow{\scriptstyle p} \\
\{E(\bullet) \xrightarrow{s} V(\bullet)\} & \xrightarrow{\ \nu\ } & \mathcal{G}
\end{array}
$$

Example 16. *As another example of the application of lifting problems to causal inference, let us consider the problem of determining whether two causal DAGs, G_1 and G_2 are Markov equivalent [48]. A key requirement here is that the immoralities of G_1 and G_2 must be the same, that is, if G_1 has a collider $A \to B \leftarrow C$, where there is no edge between A and C, then G_2 must also have the same collider, and none others. We can formulate the problem of finding colliders as the following lifting*

problem. Note that the three vertices A, B and C are bound to an actual graph instance through the category of elements ∫ δ (as was illustrated above), using the top right morphism μ. The bottom left morphism f binds these three vertices to some collider. The bottom right morphism ν requires this collider to exist in the causal graph 𝒢 with the same bindings as found by μ. The dashed morphisms h finds all solutions to this lifting problem, that is, all colliders involving the vertices A, B and C.

$$
\begin{array}{ccc}
\{A(\bullet), B(\bullet), C(\bullet)\} & \xrightarrow{\ \mu\ } & \int \delta \\
\downarrow{\scriptstyle f} & \overset{h}{\nearrow} & \downarrow{\scriptstyle p} \\
\{A(\bullet) \to B(\bullet) \leftarrow C(\bullet)\} & \xrightarrow{\ \nu\ } & \mathcal{G}
\end{array}
$$

If the category of elements is defined by a functor mapping a database schema into a table of instances, then the associated lifting problem corresponds to familiar problems like SQL queries in relational databases [4]. In our application, we can use the same machinery to formulate causal inference queries by choosing the categories appropriately. To complete the discussion, we now make the connection between universal arrows and the core notion of universal representations via the Yoneda Lemma.

6.3. Modeling Causal Interventions as Kan Extension

It is well known in category theory that ultimately every concept, from products and co-products, limits and co-limits, and ultimately even the Yoneda Lemma (see below), can be derived as special cases of the Kan extension [35]. Kan extensions intuitively are a way to approximate a functor $\mathcal{F}$ so that its domain can be extended from a category $\mathcal{C}$ to another category $\mathcal{D}$. Because it may be impossible to make commutativity work in general, Kan extensions rely on natural transformations to make the extension be the best possible approximation to $\mathcal{F}$ along $\mathcal{K}$. We want to briefly show Kan extensions can be combined with the category of elements defined above to construct causal "migration functors" that map from one causal model into another. These migration functors were originally defined in the context of database migration [4], and here we are adapting that approach to causal inference. By suitably modifying the category of elements from a set-valued functor $\delta : \mathcal{C} \to$ **Set**, to some other category, such as the category of topological spaces, namely $\delta : \mathcal{C} \to$ **Top**, we can extend the causal migration functors into solving more abstract causal inference questions. We explore the use of such constructions in the next section on Layer 4 of the UCLA hierarchy. Here, for simplicity, we restrict our focus to Kan extensions for migration functors over the category of elements defined over instances of a causal model.

Definition 27. *A* **left Kan extension** *of a functor $F : \mathcal{C} \to \mathcal{E}$ along another functor $K : \mathcal{C} \to \mathcal{D}$, is a functor $Lan_K F : \mathcal{D} \to \mathcal{E}$ with a natural transformation $\eta : F \to Lan_F \circ K$ such that for any other such pair $(G : \mathcal{D} \to \mathcal{E}, \gamma : F \to GK)$, γ factors uniquely through η. In other words, there is a unique natural transformation $\alpha : Lan_F \implies G$.*
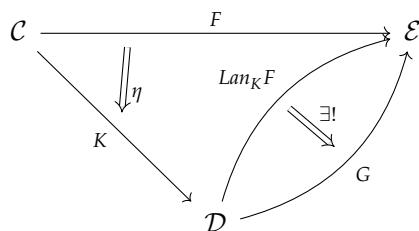
A **right Kan extension** can be defined similarly. To understand the significance of Kan extensions for causal inference, we note that under a causal intervention, when a causal category S gets modified to T, evaluating the modified causal model over a database of instances can be viewed as an example of Kan extension.

Let $\delta : S \to$ **Set** denote the original causal model defined by the category S with respect to some dataset. Let $\epsilon : T \to$ **Set** denote the effect of a causal intervention abstractly

defined as some change in the category S to T, such as deletion of an edge, as illustrated in Figure 10. Intuitively, we can consider three cases: the *pullback* Δ_F along F, which maps the effect of a causal intervention back to the original model, the *left pushforward* Σ_F and the *right pushforward* $\prod_F$, which can be seen as adjoints to the pullback Δ_F.
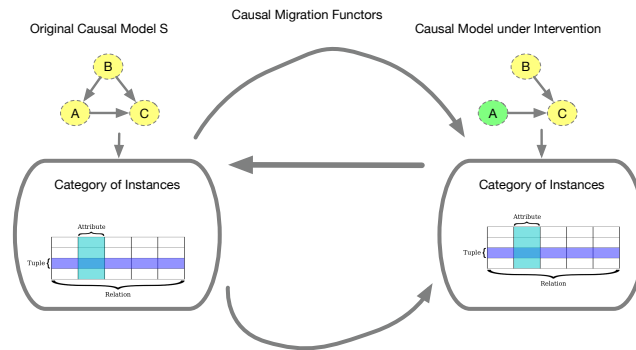

Figure 10. Kan extensions are useful in modeling the effects of a causal intervention, where in this example of a causal model over three objects A, B, and C, the object A is intervened upon, eliminating the morphism into it from object B.

Following [4], we can define three *causal migration functors* that evaluate the impact of a causal intervention with respect to a dataset of instances.

1. The functor $\Delta_F : \epsilon \to \delta$ sends the functor $\epsilon : T \to$ **Set** to the composed functor $\delta \circ F : S \to$ **Set**.
2. The functor $\Sigma_F : \delta \to \epsilon$ is the left Kan extension along F, and can be seen as the left adjoint to Δ_F.
   The functor $\prod_F : \delta \to \epsilon$ is the right Kan extension along F, and can be seen as the right adjoint to Δ_F.

To understand how to implement these functors, we use the following proposition that is stated in [4] in the context of database queries, which we are restating in the setting of causal inference.

Theorem 9. *Let $F : S \to T$ be a functor. Let $\delta : S \to$ **Set** and $\epsilon : T \to$ **Set** be two set-valued functors, which can be viewed as two instances of a causal model defined by the category S and T. If we view T as the causal category that results from a causal intervention on S (e.g., deletion of an edge), then there is a commutative diagram linking the category of elements between S and T.*

$$
\begin{array}{ccc}
\int \delta & \longrightarrow & \int \epsilon \\
\downarrow{\scriptstyle \pi_\delta} & & \downarrow{\scriptstyle \pi_\epsilon} \\
S & \xrightarrow{F} & T
\end{array}
$$

Proof. To check that the above diagram is a pullback, that is, $\int \delta \simeq S \times_T \int \delta$, or in words, the fiber product, we can check the existence of the pullback component wise by comparing the set of objects and the set of morphisms in $\int \delta$ with the respective sets in $S \times_T \int \epsilon$. $\square$

For simplicity, we defined the migration functors above with respect to an actual dataset of instances. More generally, we can compose the set-valued functor $\delta : S \to$ **Set** with a functor $\mathcal{T} :$ **Set** $\to$ **Top** to the category of topological spaces to derive a Kan extension formulation of the definition of a causal intervention. We discuss this issue in the next section on causal homotopy.

7. Layer 4 of UCLA: Causal Homotopy

Finally, we turn to discuss the role of the causal homotopy layer. To understand the reason for considering homotopy in causal inference, note that causal models can

only be determined up to some equivalence class from data, and while many causal discovery algorithms assume arbitrary interventions can be carried out, e.g., on separating sets [27], to discover the unique structure, such interventions are generally impossible to do in practical applications. The concept of *essential graph* [48] and *chain graph* [49] are attempts to formulate the notion of a "quotient space" of graphs, but similar issues arise more generally for non-graph based models as well. Thus, it is useful to understand how to formulate the notion of equivalent classes of causal models in an arbitrary category. For example, given the conditional independence structure $A \perp\!\!\!\perp B|C$, there are at least three different symmetric monoidal categorical representations that all satisfy this conditional independence [25,29,30], and we need to define the quotient space over all such equivalent categories.

In our previous work on causal homotopy [41], we exploited the connection between causal DAG graphical models and finite topological spaces [50,51]. In particular, for a DAG model $G = (V, E)$, it is possible to define a finite space topology $\mathcal{T} = (V, \mathcal{O})$, whose open sets $\mathcal{O}$ are subsets of the vertices V such that each vertex x is associated with an open set U_x defined as the intersection of all open sets that contain x. This structure is referred to an *Alexandroff* topology [52]. An intuitive way to construct an Alexandroff topology is to define the open set for each variable x by the set of its ancestors A_x, or by the set of its descendants D_x. This approach transcribes a DAG graph into a finite topological space, upon which the mathematical tools of algebraic topology can be applied to construct homotopies among equivalent causal models. Our approach below generalizes this construction to simplicial objects, as well as general categories.

7.1. The Category of Fractions: Localizing Invertible Morphisms in a Category

A principal challenge in causal discovery is that models can be inferred from data only up to an equivalence class. We can view the morphisms between equivalent causal models as "invertible" arrows, which defines a construction called an "essential" graph [48]. The problem of defining a category with a given subclass of invertible morphisms, called the category of fractions [53], is another concrete illustration of the close relationships between categories and graphs. It is also useful in the context of causal inference, as for example, in defining the Markov equivalence class of directed acyclic graphs (DAGs) as a category that is localized by considering all invertible arrows as isomorphisms. Borceux [54] has a detailed discussion of the "calculus of fractions", namely how to define a category where a subclass of morphisms are to be treated as isomorphisms. The formal definition is as follows:

Definition 28. *Consider a category $\mathcal{C}$ and a class Σ of arrows of $\mathcal{C}$. The* **category of fractions** *$\mathcal{C}(\Sigma^{-1})$ is said to exist when a category $\mathcal{C}(\Sigma^{-1})$ and a functor $\phi : \mathcal{C} \to \mathcal{C}(\Sigma^{-1})$ can be found with the following properties:*

1. *$\forall f, \phi(f)$ is an isomorphism.*
2. *If $\mathcal{D}$ is a category, and $F : \mathcal{C} \to \mathcal{D}$ is a functor such that for all morphisms $f \in \Sigma$, $F(f)$ is an isomorphism, then there exists a unique functor $G : \mathcal{C}(\Sigma^{-1}) \to \mathcal{D}$ such that $G \circ \phi = F$.*

A detailed construction of the category of fractions is given in [54], which uses the underlying directed graph skeleton associated with the category. The characterization of the Markov equivalent class of acyclic directed graphs is an example of the abstract concept of category of fractions [48]. Briefly, this condition states that two acyclic directed graphs are Markov equivalent if and only if they have the same skeleton and the same immoralities. In our previous work [41], we explored constructing homotopically invariant causal models over finite Alexandroff topological spaces, which can be seen as a special case of the UCLA framework since finite topological (Alexandroff) spaces define a category [52].

7.2. Homotopy of Simplicial Objects

We will discuss homotopy in categories more generally now. This notion of homotopy generalizes the notion of homotopy in topology, which defines why an object like a coffee cup is topologically homotopic to a doughnut (they have the same number of "holes").

Definition 29. *Let C and C' be a pair of objects in a category $\mathcal{C}$. We say C is **a retract** of C' if there exists maps $i : C \to C'$ and $r : C' \to C$ such that $r \circ i = id_C$.*

Definition 30. *Let $\mathcal{C}$ be a category. We say a morphism $f : C \to D$ is a **retract of another morphism** $f' : C \to D$ if it is a retract of f' when viewed as an object of the functor category $\textbf{Hom}([1], \mathcal{C})$. A collection of morphisms T of $\mathcal{C}$ is **closed under retracts** if for every pair of morphisms f, f' of $\mathcal{C}$, if f is a retract of f', and f' is in T, then f is also in T.*

Definition 31. *Let X and Y be simplicial sets, and suppose we are given a pair of morphisms $f_0, f_1 : X \to Y$. A **homotopy** from f_0 to f_1 is a morphism $h : \Delta^1 \times X \to Y$ satisfying $f_0 = h|_{0 \times X}$ and $f_1 = h_{1 \times X}$.*

7.3. Singular Homology

Our goal is to define an abstract notion of a causal model in terms of its underlying classifying space as a category, and show how it can be useful in defining causal homotopy. We will also clarify how it relates to determining equivalences among causal models, namely homotopical invariance, and also how it sheds light on causal identification. First, we need to define more concretely the topological n-simplex that provides a concrete way to attach a topology to a simplicial object. Our definitions below build on those given in [14]. For each integer n, define the topological space $|\Delta_n|$ realized by the object Δ_n as

$$|\Delta_n| = \{t_0, t_1, \ldots, t_n \in \mathbb{R}^{n+1} : t_0 + t_1 + \ldots + t_n = 1\}$$

This is the familiar n-dimensional simplex over n variables. For any causal model, its classifying space $|\mathcal{N}_\bullet(\mathcal{C})|$ defines a topological space. We can now define the *singular n-simplex* as a continuous mapping $\sigma : |\Delta_N| \to |\mathcal{N}_\bullet(\mathcal{C})|$. Every singular n-simplex σ induces a collection of $n-1$-dimensional simplices called *faces*, denoted as

$$d_i\sigma(t_0, \ldots, t_{n-1}) = (t_0, t_1, \ldots, t_{i-1}, 0, t_i, \ldots, t_{n-1})$$

Note that as discussed above, a causal intervention on a variable in a DAG can be modeled as applying one of these degeneracy operators d_i. The above definition shows that every such intervention has an effect on the topology associated with the causal model. Define the set of all morphisms $\text{Sing}_n(X) = \textbf{Hom}_{\textbf{Top}}(\Delta_n, |\mathcal{N}_\bullet(\mathcal{C})|)$ as the set of singular n-simplices of $|\mathcal{N}_\bullet(\mathcal{C})|$.

Definition 32. *For any topological space defined by a causal model $|\mathcal{N}_\bullet(\mathcal{C})|$, the **singular homology groups** $H_*(|\mathcal{N}_\bullet(\mathcal{C})|; \mathbf{Z})$ are defined as the homology groups of a chain complex*

$$\ldots \xrightarrow{\partial} \mathbf{Z}(Sing_2(|\mathcal{N}_\bullet(\mathcal{C})|)) \xrightarrow{\partial} \mathbf{Z}(Sing_1(|\mathcal{N}_\bullet(\mathcal{C})|)) \xrightarrow{\partial} \mathbf{Z}(Sing_0(|\mathcal{N}_\bullet(\mathcal{C})|))$$

where $\mathbf{Z}(Sing_n(|\mathcal{N}_\bullet(\mathcal{C})|))$ denotes the free Abelian group generated by the set $Sing_n(|\mathcal{N}_\bullet(\mathcal{C})|)$ and the differential ∂ is defined on the generators by the formula

$$\partial(\sigma) = \sum_{i=0}^{n} (-1)^i d_i\sigma$$

Intuitively, a chain complex builds a sequence of vector spaces that can be used to construct an algebraic invariant of a causal model from its classifying space by choosing the left **k** module **Z** to be a vector space. Each differential ∂ then becomes a linear transfor-

mation whose representation is constructed by modeling its effect on the basis elements in each $\mathbf{Z}(\mathrm{Sing}_n(X))$.

Example 17. *Let us illustrate the singular homology groups defined by an integer-valued multi-set [8] used to model conditional independence. Imsets over a DAG of three variables $N = \{a, b, c\}$ can be viewed as a finite discrete topological space. For this topological space X, the singular homology groups $H_*(X; \mathbf{Z})$ are defined as the homology groups of a chain complex*

$$\mathbf{Z}(Sing_3(X)) \xrightarrow{\partial} \mathbf{Z}(Sing_2(X)) \xrightarrow{\partial} \mathbf{Z}(Sing_1(X)) \xrightarrow{\partial} \mathbf{Z}(Sing_0(X))$$

where $\mathbf{Z}(Sing_i(X))$ denotes the free Abelian group generated by the set $Sing_i(X)$ and the differential ∂ is defined on the generators by the formula

$$\partial(\sigma) = \sum_{i=0}^{4} (-1)^i d_i \sigma$$

The set $Sing_n(X)$ is the set of all morphisms $\mathbf{Hom}_{Top}(|\Delta_n|, X)$. For an imset over the three variables $N = \{a, b, c\}$, we can define the singular n-simplex σ as:

$$\sigma : |\Delta^4| \to X \;\; where \;\; |\Delta^n| = \{t_0, t_1, t_2, t_3 \in [0,1]^4 : t_0 + t_1 + t_2 + t_3 = 1\}$$

The n-simplex σ has a collection of faces denoted as $d_0\sigma, d_1\sigma, d_2\sigma$ and $d_3\sigma$. If we pick the k-left module $\mathbf{Z}$ as the vector space over real numbers $\mathbb{R}$, then the above chain complex represents a sequence of vector spaces that can be used to construct an algebraic invariant of a topological space defined by the integer-valued multiset. Each differential ∂ then becomes a linear transformation whose representation is constructed by modeling its effect on the basis elements in each $\mathbf{Z}(Sing_n(X))$. An alternate approach to constructing a chain homology for an integer-valued multiset is to use Möbius inversion to define the chain complex in terms of the nerve of a category (see our recent work on categoroids [15] for details).

7.4. Classifying Spaces and Homotopy Colimits

Building on the intuition proposed above, we now introduce a formal way to define causal effects in our framework, which relies on the construction of a topological space associated with the nerve of a category. As we saw above, the nerve of a category is a full and faithful embedding of a category as a simplicial object.

Definition 33. *The **classifying space** of a causal model defined as a category $\mathcal{C}$ is the topological space associated with the nerve of the category $|N_\bullet \mathcal{C}|$.*

To understand the classifying space $|N_\bullet \mathcal{C}|$ of a causal model defined as a category $\mathcal{C}$, let us go over some simple examples to gain some insight.

Example 18. *For any set X, which can be defined as a discrete category $\mathcal{C}_X$ with no non-trivial morphisms, the classifying space $|N_\bullet \mathcal{C}_X|$ is just the discrete topology over X (where the open sets are all possible subsets of X).*

Example 19. *If we take a causal model defined as a partially ordered set $[n]$, with its usual order-preserving morphisms, then the nerve of $[n]$ is isomorphic to the representable functor $\delta(-, [n])$, as shown by the Yoneda Lemma, and in that case, the classifying space is just the topological space Δ_n defined above.*

Example 20. *In our earlier work on causal homotopy [41], we associated with any finite causal DAG $\mathcal{G}$, a finite Alexandroff topological space, where the open sets of the topology corresponding to the* down *sets or* upsets *of descendants or ancestors, respectively. Since any causal DAG model $\mathcal{G}$*

induces a partial ordering, we can then define the classifying space of a causal DAG in terms of the topological space associated with the nerve of the DAG, namely $|N_\bullet \mathcal{G}|$.

Example 21. *Witsenhausen [6] defined a measure-theoretic notion of causality called the* intrinsic model. *An intrinsic model $\mathcal{M} = (\alpha, U_\alpha, \mathcal{I}_\alpha)_{\alpha \in A}$, where the parameters of the intrinsic causal model over n variables A are defined in terms of a collection of measurable functions over each variable's information field $\mathcal{I}_\alpha$ (a subfield of the product σ-algebra over all variables upon which it depends), where U_α is the space over which α takes its values. Heymann et al. [7] showed recently that Witsenhausen's intrinsic model generalizes Pearl's d-separation condition, and can be used to define a rich set of causal models that includes cycles and feedback, as well as more refined notions of conditional d-separation. The definition of causality in an intrinsic model is based on structuring the information fields of every variable in such a way that it is possible to sequentially order them for any particular instance of the underlying sample space. It is possible to define a topology on the underlying variables (which Witsenhausen referred to as* agents*), by defining* subsystem *of variables $B \subseteq A$ such that every variable $\alpha \in B$ has an information field that only depends on the information fields of members in its subset B, that is $\forall \alpha \in B$, the condition states that $\mathcal{I}_\alpha \subseteq \mathcal{F}_B$, where $\mathcal{F}_B$ is the induced product information field over the subset of variables B. Witsenhausen proves that the collection of subsystems forms a finite topology on A. We can then define the classifying space of an intrinsic causal model to be the topological space associated with the nerve of an intrinsic model $\mathcal{M}$, namely $|N_\bullet \mathcal{M}|$.*

We now want to bring in the set-valued functor mapping each causal category $\mathcal{C}$ to the actual experiment used, e.g., in a clinical trial [9], to evaluate average treatment effect or quantify the effect of a **do** calculus intervention [5] We can then compute the topological space prior to intervention, and subsequent to intervention, and compare the two topological spaces in terms of their algebraic invariants (e.g., the chain complex, as described below).

Definition 34. *The* **homotopy colimit** *of a causal model defined as nerve of the category of elements associated with the set-valued functor $\delta : \mathcal{C} \to$ **Set** mapping the causal category $\mathcal{C}$ to a dataset, namely $N_\bullet(\int \delta)$.*

In general, we may want to evaluate the homotopy colimit of a causal model not only with respect to the data used in a causal experiment, but also with respect to some underlying topological space or some measurable space. We can extend the above definition straightforwardly to these cases using an appropriate functor $\mathcal{T}: $ **Set** $\to$ **Top**, or alternatively $\mathcal{M}: $ **Set** $\to$ **Meas**. These augmented constructions can then be defined with respect to a more general notion called the *homotopy colimit* [40] of a causal model.

Definition 35. *The* **topological homotopy colimit** *$hocolim_{\mathcal{T} \circ \delta}$ of a causal model associated with a category $\mathcal{C}$, along with its associated category of elements associated with a set-valued functor $\delta : \mathcal{C} \to$ **Set**, and a topological functor $\mathcal{T} : $ **Set** $\to$ **Top** is isomorphic to topological space associated with the nerve of the category of elements, that is $hocolim_{\mathcal{T} \circ \delta} \simeq |N_\bullet(\int \delta)|$.*

Example 22. *The classifying space $|N_\bullet \mathcal{C}_{CDU}|$ associated with CDU symmetric monoidal category encoding of a causal Bayesian DAG is defined using the monoidal category $(\mathbf{C}, \otimes, I)$, where each object A has a copy map $C_A : A \to A \otimes A$, and discarding map $D_A : A \to I$, and a uniform state map $U_A : I \to A$, is defined as the topological realization of its nerve. As before, the nerve $N_n(\mathcal{C})$ of the CDU category is defined as the set of sequences of composable morphisms of length n.*

$$\{C_o \xrightarrow{f_1} C_1 \xrightarrow{f_2} \ldots \xrightarrow{f_n} C_n \mid C_i \text{ is an object in } \mathcal{C}, f_i \text{ is a morphism in } \mathcal{C}\}$$

Note that the CDU category was associated with a CDU functor $F : \mathbf{Syn}_G \to \mathbf{Stoch}$ to the category of stochastic matrices. We can now define the homotopy colimit $hocolim_{\mathcal{F}}$ of the CDU causal model associated with the CDU category $\mathcal{C}$, along with its associated category of elements

associated with a set-valued functor $\delta : \mathcal{C} \to$ **Set**, *and a topological functor* $\mathcal{F}:$ **Set** $\to$ **Stoch** *is isomorphic to topological space associated with the nerve of the category of elements over the composed functor, that is* $\mathrm{hocolim}_{\mathcal{F} \circ \delta}$.

7.5. Defining Causal Effect

Finally, we turn to defining causal effect using the notion of classifying space and homotopy colimits, as defined above. Space does not permit a complete discussion of this topic, but the basic idea is that once a causal model is defined as a topological space, there are a large number of ways of comparing two topological spaces from analyzing their chain complexes, or using a topological data analysis method such as UMAP [2].

Definition 36. *Let the classifying space under "treatment" be defined as the topological space* $|N_\bullet \mathcal{C}_1|$ *associated with the nerve of category* $\mathcal{C}_1$ *under some intervention, which may result in a topological deformation of the model (e.g., deletion of an edge). Similarly, the classifying space under "no treatment" be defined as the* $|N_\bullet \mathcal{C}_0|$ *under a no-treatment setting, with no intervention. A* **causally non-isomorphic effect** *exists between categories* $\mathcal{C}_1$ *and* $\mathcal{C}_0$, *or* $\mathcal{C}_1 \ncong \mathcal{C}_0$ *if and only if there is no invertible morphism* $f : |N_\bullet \mathcal{C}_1| \to N_\bullet (\mathcal{C}_0|$ *between the "treatment" and "no-treatment" topological spaces, namely* f *must be both* left invertible *and* right invertible.

There is an equivalent notion of causal effect using the homotopy colimit definition proposed above, which defines the nerve functor using the category of elements. This version is particularly useful in the context of evaluating a causal model over a dataset.

Definition 37. *Let the homotopy colimit* $\mathrm{hocolim}_1 = |N_\bullet(\int \delta_1)|$ *be the topological space associated with a causal category* $\mathcal{C}_1$ *under the "treatment' condition be defined with respect to an associated category of elements defined by a set-valued functor* $\delta_1 : \mathcal{C} \to$ **Set** *over a dataset of "treated" variables, and corresponding "no-treatment"* $\mathrm{hocolim}_0 = |N_\bullet(\int \delta_0)|$ *be the topological space of a causal model associated with a category* $\mathcal{C}_0$ *be defined over an associated category of elements defined by a set-valued functor* $\delta_0 : \mathcal{C} \to$ **Set** *over a dataset of "placebo" variables. A* **causally non-isomorphic effect** *exists between categories* $\mathcal{C}_1$ *and* $\mathcal{C}_0$, *or* $\mathcal{C}_1 \ncong \mathcal{C}_0$ *if and only if there is no invertible morphism* $f : |N_\bullet(\int \delta_1)| \to |N_\bullet(\delta_0)|$ *between the "treatment" and "no-treatment" homotopy colimit topological spaces, namely* f *must be both* left invertible *and* right invertible.

We can define an equivalent "do-calculus" like version of the causal effect definitions above for the case when a causal model defined as a graph structure is manipulated by an intervention that deletes an edge, or does some more sophisticated type of "category" surgery.

8. Contributions of Our Paper

We summarize the principal contributions of our paper. Our principal contribution is the development of the notion of "universal causality", a representation-independent definition whose goal is to elucidate the "universal' properties of causal inference. Our work is inspired by other work, for example separoids [10] elucidates the concept of conditional independence in a representation-independent manner, which applies to conditional independence in probability theory, statistics, and geometry. Another example is the concept of *Grothendieck topology* [36], which defines topology abstractly in the context of any category. Implicit in these constructions is the abstraction of a specific construct—conditional independence or topology–in a manner that lets it be studied across a wide range of representations. Similarly, UC is intended to be an abstract characterization of causality.

1. **Universal Arrow:** We used universal arrows as a unifying principle in UC, which allows synchronizing causal changes at different levels of the UCLA hierarchy. Universal arrows set up a correspondence between a "forgetful" functor and its left adjoint"free" functor. In the application to causal inference, universal arrows, for example, define forgetful and free functors between the category of conditional in-

dependence structures, such as separoids, from the category of actual causal models (e.g., as symmetric monoidal categories of causal DAG models [25,29,30]).

2. **Causal reproducing property:** The universal arrow property leads to the powerful Yoneda Lemma, which provides the foundational result embodied in the causal reproducing property. The CRP implies that all causal influences between two objects X and Y in a category $\mathcal{C}$ are representable in the functor category of presheaves, namely

$$\mathbf{Hom}_{\mathcal{C}}(X, Y) \simeq \mathbf{Nat}(\mathbf{Hom}_{\mathcal{C}}(-, X), \mathbf{Hom}_{\mathcal{C}}(-, Y))$$

3. **Causal interventions as a higher-order category:** Most causal discovery algorithms require a sequence of interventions, which naturally compose to form a category. We introduced the framework of higher-order category theory using simplicial sets and objects to define a category over causal interventions. Simplicial objects provides an elegant and general way of extracting parts of a compositional structure, and its associated lifting problems define when a partial fragment of a causal model can be "put back" together into a complete model.

4. **Nerve of a causal model:** We used the nerve construction to set up a functor between a casual category and its associated simplicial object, which is a fully faithful embedding of any category as a simplicial object. Its left adjoint functor, which maps a simplicial set into a category, is a lossy representation that only preserves structure up to $n \leq 2$ simplices. Simplicial sets suggest a way to define higher-order causal models, a topic for future work.

5. **Relational causal models:** The Grothendieck category of elements is closely related to the notion of *ground graphs* in relational causal models [32], which gives a rich source of applications of causal inference. Any relational database defines a category [4], and our paper shows how to formulate causal inference in the rich space of relational enterprise datasets.

6. **Lifting Problem:** Associated with each pair of layers of the UCLA hierarchy is a lifting problem over a suitable category of elements, from simplicial category of elements, to a category of elements over a dataset, to a category of elements over a topological space. In general, the Grothendieck category of elements is a way to embed each object in a category into the category of all categories **Cat**. This construction has many elegant properties, which deserves further exploration in a subsequent paper.

7. **Homotopy colimits and Classifying Spaces:** We defined causal effect in terms of the classifying space associated with the nerve of a causal category, and with the homotopy colimit of the nerve of the category of elements. These structures have been extensively explored in the study of homotopy in category theory [40], and there are many advanced techniques that can be brought to bear on this problem, such as *model categories* [55].

9. Future Work

There are many directions for future work, and we summarize a few of them below.

1. **Higher-order causality:** Our use of simplicial sets and objects suggests a way of defining higher-order causality, as simplicial sets generalize directed graphs, categories, and partial orders. Simplicial sets permit modeling the interaction between groups of objects, which naturally applies to cases of causal inference with *interference*, where the stable unit treatment value assumption (SUTVA) [9] is violated. Zigler and Papadogeorgou [33] explores an application to causal interference, where the treatment units (e.g., power plants) and response units (e.g., people living close to power plants) have a complex set of interactions, where a particular treatment may affect many individuals. These types of problems can be studied using higher-order degeneracy operators over oriented n-simplexes.

2. **Causal Discovery from Conditional Independence Oracles:** The problem of causal discovery can then be rigorously formulated as a lifting problem as well, where the

conditional independence oracle is defined as a solution to a lifting problem. More specifically, it is possible to define a Grothendieck category of elements for a functor F: **Graph** → **Separoids** mapping the category of directed graphs into the category of separoids, which define its equivalent set of conditional independence statements. The Grothendieck fibration in this case maps the category of elements, combining conditional independence properties and graph objects, into the category **Graph**. Algorithms proposed in the literature, such as [27], can be seen as queries in a lifting problem, analogous to the lifting problems defined for the UCLA hierarchy. This approach can be extended to causal discovery over higher-order categories.

3. **Grothendieck Topology:** Analogous to the representation-independent definition of conditional independence using separoids, our longstanding goal has been to define causality purely in terms of a categorical structure. The Grothendieck topology $\mathcal{J}$ for any category, which leads to the concept of a *site* [36], is defined such that for any object c in $\mathcal{C}$, a *sieve S* is a family of morphisms, all with co-domain c such that

$$f \in S \to f \circ g \in S$$

for any g where the composition is defined. A Grothendieck topology $\mathcal{J}$ on category $\mathcal{C}$ defines a sieve $J(c)$ for each object c such that the following properties hold: (i) the maximal sieve $t_c = \{f | cod(f) = c\}$ is in $J(c)$. There is an additional stability condition and a transitive closure condition. An interesting problem for future work is to define causal inference over sheaves of a site, using the concept of Grothendieck topologies. Any causal intervention that, for instance, deletes an edge, would change the Grothendieck topology embodied in the structure of sieves.

4. **Gröbner Causal Models:** Another direction for future work is to construct Gröbner representations of causal categories. Sam and Snowden [56] define a general **Gröbner** representation for combinatorial categories, which apply to causal models as well. Specifically, denote $\mathbf{Rep}_{\mathbf{k}}(\mathcal{C})$ as the category of representations of a causal model $\mathcal{C}$, where **k** is a non-zero ring, and $\mathrm{Mod}_{\mathbf{k}}$ is the category of left-**k** modules. Thus, we can define a **representation** of a causal category $\mathcal{C}$ as a functor $\mathcal{C} \to \mathrm{Mod}_{\mathbf{k}}$. Let x be an object of $\mathcal{C}$. Define a representation P_x of $\mathcal{C}$ as a left **k**-module, where $P_x(y) = \mathbf{k}[\mathbf{Hom}_{\mathcal{C}}(x,y)]$, that is, $P_x(y)$ is the free left **k**-module with basis $\mathbf{Hom}_{\mathcal{C}}(x,y)$. For any particular morphism $f : x \to y$, let e_f denote the corresponding element of $P_x(y)$. Broadly speaking, this approach generalizes the work on modeling graphical models as algebraic varieties [20,57,58], and ideals on partially ordered sets (posets) [59]. The intuitive idea is that a representation of a category can be defined as an abstract Gröbner basis over an ideal defined on a module whose basis is defined using the free algebra generated by the set of all morphisms out of an object. This approach provides an alternative way of parameterizing causal models defined as combinatorial categories.

10. Summary

In this paper, we proposed a framework called Universal Causality (UC) for causal inference using the tools of category theory. Specifically, we described a layered hierarchical architecture called UCLA (Universal Causality Layered Architecture), where causal inference is modeled at multiple levels of categorical abstraction. At the top-most level, causal inference is modeled using a higher-order category of simplicial sets and objects, defined as contravariant functors from the category of ordinal numbers Δ, whose objects are the ordered natural numbers $[n] = \{0, \ldots, n\}$, and whose morphisms are order-preserving injections and surjections. Causal "surgery" is then modeled as the action of a contravariant functor from the category Δ into a causal model. At the second layer, causal models are defined by a category consisting of a collection of objects, such as the entities in a relational database, and morphisms between objects can be viewed as attributes relating entities. The third categorical abstract layer corresponds to the data layer in causal inference, where each causal object is mapped into a set of instances, modeled using the category of sets and

morphisms are functions between sets. The fourth layer comprises of additional structure imposed on the instance layer above, such as a topological space, a measurable space or a probability space, or more generally, a locale. Between every pair of layers in UCLA are functors that map objects and morphisms from the domain category to the co-domain category. Each functor between layers is characterized by a universal arrow, which defines an isomorphism between every pair of categorical layers. These universal arrows define universal elements and representations through the Yoneda Lemma, and in turn lead to a new category of elements based on a construction introduced by Grothendieck. Causal inference between each pair of layers is defined as a lifting problem, a commutative diagram whose objects are categories, and whose morphisms are functors that are characterized as different types of fibrations. We defined causal effect in the UCLA framework using the notion of homotopy colimits associated with the nerve of a category. We illustrate the UCLA architecture using a diverse set of examples.

References

1. Fong, B.; Spivak, D.I. *Seven Sketches in Compositionality: An Invitation to Applied Category Theory*; Cambridge University Press: Cambridge, UK, 2018.
2. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**, arXiv:1802.03426.
3. Carlsson, G.E.; Mémoli, F. Classifying Clustering Schemes. *Found. Comput. Math.* **2013**, *13*, 221–252. [CrossRef]
4. Spivak, D.I. Database queries and constraints via lifting problems. *Math. Struct. Comput. Sci.* **2013**, *24*, e240602. [CrossRef]
5. Pearl, J. *Causality: Models, Reasoning and Inference*, 2nd ed.; Cambridge University Press: Cambridge, MA, USA, 2009.
6. Witsenhausen, H.S. The Intrinsic Model for Discrete Stochastic Control: Some Open Problems. In Proceedings of the Control Theory, Numerical Methods and Computer Systems Modelling, Rocquencourt, France, 17–21 June1974; Bensoussan, A., Lions, J.L., Eds.; Springer: Berlin/Heidelberg, Germany, 1975; pp. 322–335.
7. Heymann, B.; de Lara, M.; Chancelier, J.P. Causal Information with Information Fields. In Proceedings of the Neural Inforamtion Processing Systems Workshop on Causal Discovery and Causality-Inspired Machine Learning, Online, 11 December 2020.
8. Studeny, M. *Probabilistic Conditional Independence Structures*; Information Science and Statistics; Springer: London, UK, 2010.
9. Imbens, G.W.; Rubin, D.B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*; Cambridge University Press: Cambridge, MA, USA, 2015.
10. Dawid, A.P. Separoids: A Mathematical Framework for Conditional Independence and Irrelevance. *Ann. Math. Artif. Intell.* **2001**, *32*, 335–372. [CrossRef]
11. Lewis, D. Counterfactuals and comparative possibility. *J. Philos. Log.* **1973**, *2*, 418–446. [CrossRef]
12. May, J. *Simplicial Objects in Algebraic Topology*; University of Chicago Press: Chicago, IL, USA, 1992.
13. Joyal, A. Quasi-categories and Kan complexes. *J. Pure Appl. Algebra* **2002**, *175*, 207–222. [CrossRef]
14. Lurie, J. *Higher Topos Theory*; Annals of mathematics studies; Princeton University Press: Princeton, NJ, USA, 2009.
15. Mahadevan, S. Categoroids: Universal Conditional Independence. *arXiv* **2022**, arXiv:2208.11077.
16. Pearl, J. *Probabilistic Reasoning in Intelligent Systems—Networks of Plausible Inference*; Morgan Kaufmann Series in Representation and Reasoning; Morgan Kaufmann: San Francisco, CA, USA, 1989.
17. Gavrilovich, M. The unreasonable power of the lifting property in elementary mathematics. *arXiv* **2017**, arXiv:1707.06615.
18. Darwin, C. *The Effect of Cross and Self-Fertilization in the Vegetable Kingdom*; John Murray: London, UK, 1876.
19. Wright, S. Correlation and causation. *J. Agric. Res.* **1921**, *20*, 557–585.
20. Beerenwinkel, N.; Eriksson, N.; Sturmfels, B. Evolution on distributive lattices. *J. Theor. Biol.* **2006**, *242*, 409–420. [CrossRef]
21. Maier, M.; Marazopoulou, K.; Arbour, D.; Jensen, D. A Sound and Complete Algorithm for Learning Causal Models from Relational Data. *arXiv* **2013**, arXiv:1309.6843. https://doi.org/10.48550/ARXIV.1309.6843.
22. Mahadevan, S. Universal Decision Models. *arXiv* **2021**, arXiv:2110.15431.
23. Nagurney, A. *Network Economics: A Variational Inequality Approach*; Kluwer Academic Press: Dordrecht, The Netherlands; Boston, MA, USA, 1999.
24. Hansen, J.; Ghrist, R. Opinion Dynamics on Discourse Sheaves. *arXiv* **2020**, arXiv:2005.12798.
25. Jacobs, B.; Kissinger, A.; Zanasi, F. Causal Inference by String Diagram Surgery. *arXiv* **2018**, arXiv:1811.08338.
26. Muandet, K.; Kanagawa, M.; Saengkyongam, S.; Marukatat, S. Counterfactual Mean Embeddings. *J. Mach. Learn. Res.* **2021**, *22*, 1–71.

27. Kocaoglu, M.; Shanmugam, K.; Bareinboim, E. Experimental Design for Learning Causal Graphs with Latent Variables. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; pp. 7018–7028.

28. Addanki, R.; Kasiviswanathan, S.P.; McGregor, A.; Musco, C. Efficient Intervention Design for Causal Discovery with Latents. *arXiv* **2020**, arXiv:2005.11736.

29. Fong, B. Causal Theories: A Categorical Perspective on Bayesian Networks. *arXiv* **2012**, arXiv:1301.6201.

30. Fritz, T.; Klingler, A. The d-Separation Criterion in Categorical Probability. *J. Mach. Learn. Res.* **2023**, *24*, 1–49.

31. Mahadevan, S. Unifying Causal Inference and Reinforcement Learning Using Higher-Order Category Theory. *arXiv* **2022**, arXiv:2209.06262.

32. Ahsan, R.; Arbour, D.; Zheleva, E. Relational Causal Models with Cycles:Representation and Reasoning. *arXiv* **2022**, arXiv:2202.10706.

33. Zigler, C.M.; Papadogeorgou, G. Bipartite Causal Inference with Interference. *arXiv* **2018**, arXiv:1807.08660.

34. Lauritzen, S. *Graphical Models*; Oxford University Press: Oxford, UK, 1996.

35. MacLane, S. *Categories for the Working Mathematician*; Springer: New York, NY, USA, 1971; Volume 5, p. ix+262.

36. MacLane, S.; Moerdijk, I. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*; Springer: Berlin/Heidelberg, Germany, 1994.

37. Riehl, E. *Category Theory in Context*; Aurora: Dover Modern Math Originals; Dover Publications: New York, NY, USA, 2017.

38. Joyal, A.; Nielsen, M.; Winskel, G. Bisimulation from Open Maps. *Inf. Comput.* **1996**, *127*, 164–185. .: 10.1006/inco.1996.0057. [CrossRef]

39. May, J. *A Concise Course in Algebraic Topology*; Chicago Lectures in Mathematics, University of Chicago Press: London, UK, 1999.

40. Richter, B. *From Categories to Homotopy Theory*; Cambridge Studies in Advanced Mathematics, Cambridge University Press: Cambridge, UK, 2020.

41. Mahadevan, S. Causal Homotopy. *arXiv* **2021**, arXiv:2112.01847. [CrossRef]

42. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.

43. Boardman, M.; Vogt, R. *Homotopy Invariant Algebraic Structures on Topological Spaces*; Springer: Berlin/Heidelberg, Germany, 1973.

44. Janzing, D.; Balduzzi, D.; Grosse-Wentrup, M.; Schölkopf, B. Quantifying causal influences. *Ann. Stat.* **2013**, *41*, 2324–2358. [CrossRef]

45. Coecke, B.; Fritz, T.; Spekkens, R.W. A mathematical theory of resources. *Inf. Comput.* **2016**, *250*, 59–86. [CrossRef]

46. Spivak, D.I. Simplicial Databases. *arXiv* **2009**, arXiv:0904.2012.

47. Spivak, D.I.; Kent, R.E. Ologs: A Categorical Framework for Knowledge Representation. *PLoS ONE* **2012**, *7*, e24274. [CrossRef]

48. Andersson, S.A.; Madigan, D.; Perlman, M.D. A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Stat.* **1997**, *25*, 505–541. [CrossRef]

49. Lauritzen, S.L.; Richardson, T.S. Chain graph models and their causal interpretations. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 321–348. [CrossRef]

50. Barmak, J.A. *Algebraic Topology of Finite Topological Spaces and Applications*; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 2011.

51. Stong, R.E. Finite topological spaces. *Trans. Amer. Math. Soc.* **1966**, *123*, 325–340. [CrossRef]

52. Alexandroff, P.S. *Combinatorial Topology*; Graylock Press: Toronto, ON, Canada, 1956; Volume 1.

53. Gabriel, P.; Gabriel, P.; Zisman, M. *Calculus of Fractions and Homotopy Theory*; Calculus of Fractions and Homotopy Theory; Springer: New York, NY, USA, 1967.

54. Borceux, F. *Handbook of Categorical Algebra*; Encyclopedia of Mathematics and Its Applications; Cambridge University Press: New York, NY, USA, 1994; Volume 1.

55. Quillen, D.G. *Homotopical Algebra*; Springer: Berlin/Heidelberg, Germany, 1967.

56. Sam, S.; Snowden, A. Gröbner methods for representations of combinatorial categories. *J. Am. Math. Soc.* **2016**, *30*, 159–203. [CrossRef]

57. Beerenwinkel, N.; Eriksson, N.; Sturmfels, B. Conjunctive Bayesian networks. *Bernoulli* **2007**, *13*, 893–909. [CrossRef]

58. Geiger, D.; Meek, C.; Sturmfels, B. On the toric algebra of graphical models. *Ann. Stat.* **2006**, *34*, 1463–1492. [CrossRef]

59. Hibi, T. Distributive lattices, affine semigroup rings and algebras with straightening laws. *Commut. Algebra Comb.* **1987**, *11*, 93–109.