Object Conditioning for Causal Inference

David Jensen
jensen@cs.umass.edu

Javier Burroni
jburroni@cs.umass.edu

Matthew Rattigan
rattigan@cs.umass.edu

College of Information and Computer Sciences
University of Massachusetts Amherst
Amherst, MA 01003-9264

Abstract

We describe and analyze a form of conditioning that is widely applied within social science and applied statistics but that is virtually unknown within causal graphical models. This approach, which we term *object conditioning*, can adjust for the effects of latent confounders and yet avoid the pitfall of conditioning on colliders. We describe object conditioning using plate models and show how its probabilistic implications can be explained using the property of exchangeability. We show that several seemingly obvious interpretations of object conditioning are insufficient to describe its probabilistic implications. Finally, we use object conditioning to describe and unify key aspects of a diverse set of techniques for causal inference, including within-subjects designs, difference-in-differences designs, and interrupted time-series designs.

1 INTRODUCTION

A substantial theoretical and methodological infrastructure has been developed to facilitate accurate estimates of causal dependence from observational data. Two dominant and almost entirely compatible frameworks have been developed based on causal graphical models (Pearl, 2009; Spirtes *et al.*, 2000) and potential outcomes (Rubin, 2005; Imbens and Rubin, 2015). Both frameworks rely heavily on the statistical concept of *conditioning*, particularly to adjust causal estimates for the effects of potential confounding variables. Several methodological innovations facilitate forms of conditioning, including propensity score analysis (Rosenbaum and Rubin, 1983), doubly robust estimators (Bang and Robins, 2005), and non-parametric estimators (Hill, 2011; Athey and Imbens, 2016).

The accuracy of causal estimates based on conditioning rests on several assumptions that can be difficult to meet in practice. Accurate causal estimates often require that all confounding variables have been correctly measured and that their effects have been correctly modeled. The relatively low probability that these assumptions can be simultaneously satisfied has motivated interest in methods that partially or entirely escape these assumptions. Some of these methods, such as the FCI algorithm and its successors (Spirtes *et al.*, 2000; Colombo *et al.*, 2012), have been developed primarily within the causal graphical models community. Others, such as instrumental variable designs (Angrist *et al.*, 1996), were originally developed within fields most commonly associated with the potential outcomes framework, although these methods have been widely adopted, analyzed, and extended within the framework of causal graphical models (e.g., Pearl, 1995, 2009; Sharma *et al.*, 2016; Sharma, 2017).

Another large family of methods that partially escapes these assumptions has been widely adopted within communities that use the potential outcomes framework and yet has received almost no attention within communities that predominantly employ causal graphical models. This family of methods includes the closely related methods of within-subject designs (Greenwald, 1976), multilevel models (Goldstein, 1987, 2010), interrupted time-series designs (Shadish *et al.*, 2001), and difference-in-differences designs (Shadish *et al.*, 2001). The number of papers that reference *each member* of this family of methods is of the same order of magnitude as the *entire* literature that references causal graphical models, yet this family of methods has received almost no attention within the literature on causal graphical models.

This paper directly addresses that gap by defining and analyzing *object conditioning*, the central strategy employed by each method in this family. We use the syntax and semantics of plate models to formally describe object conditioning. We show that, under fairly mild assumptions, object conditioning confers a surprising and useful set of benefits. It can provide accurate causal estimates, even when some confounders are measured poorly or not at all; it provides these estimates without risk of collider bias; and it can reduce variability due to additional covariates that are causes of either treatment or outcome alone. We show that several plausible theories drawn from within the current framework of causal graphical models cannot adequately explain the effects of object conditioning. Finally, we show how object conditioning describes and unifies several widely used techniques.

2 EXAMPLE

Consider the challenge of causal inference about a complex social media system such as Stack Overflow, a popular question-and-answer website devoted to the topic of computer programming. The site makes nearly all data it generates publicly available, including detailed information about users, questions, answers, badges, and topics. Such data make it possible to examine causal claims about how the design and incentives of the site influence user behavior (e.g., Oktay *et al.*, 2010; Marder, 2015).

For example, we might conjecture that the length of a posted question might be one cause of the number of answers the question receives. We refer to length as a *treatment* and answer count as an *outcome*. A simple analysis would merely examine the joint distribution of length and answer count for a large number of questions. However, a savvy analyst would also consider other factors that could cause treatment, outcome, or both. For example, some novice programmers may be unskilled at writing concise questions and may also tend to ask questions that draw many potential respondents due to their simple content (rather than length). This situation could be portrayed graphically by the generic model shown in Figure 1a, where the variable X is length, Y is answer count, Z is programming experience, and the plate C represents questions and the plate P represents users.

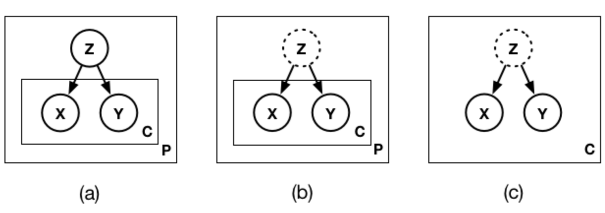

Figure 1: Three simple plate models with child (C) and parent (P) plates. (a) A multi-object unit with the variable Z measured. (b) A multi-object unit with the variable Z latent. (c) A single-object unit with the variable Z latent.

Ideally, we would be able to measure all potential confounding factors Z and condition on their effects. However, many such factors are notoriously difficult to measure quantitatively and reliably. Fortunately, such measurement can be avoided by using a more nuanced analysis that employs a *within-subjects design*, in which the length and answer-count of multiple postings *from the same individual* are compared and then the within-subject comparisons are aggregated over a large number of such individuals. In principle, this approach can eliminate the effects of many potential latent confounders simultaneously because the value of these confounders can reasonably be assumed to be constant within subjects.

A within-subjects design is a method for causal inference that employs object conditioning, the topic of this paper. Defined more formally in Section 3, this approach conditions on the *identities of individuals* (e.g., *User=Robert456*) rather than the more conventional choice of conditioning on the *values of variables* of those individuals (e.g., *programming-ability=low*). Object conditioning does not rule out conditioning on variables, but merely adds to the array of tools available for causal analysis.

Object conditioning has substantial value for estimating causal dependence. It can reduce variance in causal estimates by adjusting for other causes of treatment or outcome (regardless of whether those variables are observed or latent). More importantly, object conditioning can reduce bias in causal estimates by eliminating the effects of potential confounders that may induce statistical dependence between treatment and outcome (again, regardless of whether those variables are observed or latent).

A simple (and, as we show later, incorrect) view of object conditioning is that it simultaneously conditions on all variables of an object (e.g., a Stack Overflow user). Under this simple view, object conditioning would entail substantial risks because it could induce dependence between treatment-outcome pairs when one or more of the variables of user represents a collider rather than a confounder. That is, if treatment and outcome simultaneously cause a variable of user, then object conditioning on user could induce dependence between treatment and outcome and *increase* the bias of our causal estimate. This effect is referred to as Berkson's bias or collider bias (Elwert and Winship, 2014). This prospect is particularly disturbing because object conditioning provides no fine-grained choice about which variables are in the apparent conditioning set. Instead, object conditioning would appear to condition on a nearly infinite number of possible variables, some of which might be colliders.

Surprisingly, evidence exists that object conditioning achieves the advantages noted above (removing the effects of latent variables) *without* incurring the potential risks (collider bias). Specifically, simulation results from

Rattigan *et al.* (2011) indicate that when a variable Z of a parent object causes two variables X and Y of a child object (a confounder), object conditioning on the parent object appears to produce the same conditional independence implications as conditioning on the parent variable Z, even when Z is latent. In addition, when Z *is caused by* X and Y (a collider), object conditioning on the parent object appears to have directly the opposite conditional independence implications as conditioning on Z. That is, object conditioning does not appear to induce dependence between X and Y in such situations.

In the following sections, we ask and answer a set of fundamental research questions about object conditioning and its effects: (1) What are the probabilistic implications of object conditioning? (2) Can the effects of object conditioning be explained by known principles of directed graphical models? (3) What new principles, if any, are necessary to correctly represent the effects of object conditioning?

3 OBJECT CONDITIONING

Object conditioning can be formally defined within the framework of plate models, a frequently used formalism in graphical models. We review this formalism, define object conditioning, discuss a key probabilistic concept necessary to understand object conditioning, and discuss the probabilistic implications of object conditioning.

3.1 PLATE MODELS

Plate models are a common representation for complex forms of probabilistic graphical models (Buntine, 1994; Koller and Friedman, 2009). Figure 1 shows three simple plate models. More general and expressive representations for such models exist, but plate models are still widely used because of their simple and familiar notation. Following Koller & Friedman (2007), we briefly review key aspects of plate models below.

Objects are a basic element of plate models represented in the graphical notation as rectangles referred to as *plates*. Objects typically correspond to entities that have physical or conceptual existence in common parlance. Object instances within a domain of analysis can be divided into a set of mutually exclusive and collectively exhaustive classes $O = \{O_1 \dots O_k\}$ which define their type. Examples of object classes mentioned in Section 2 include users, questions, answers, badges, and topics. Throughout this paper, we use letters in the lower right corner of plates to denote the object type denoted by that plate (e.g., the letters C and P in Figure 1) whereas some plate notation uses similar letters to denote the cardinality of the corresponding objects. When needed, we use cardinality notation (e.g., $|C|$) to denote the latter quantity.

Relationships between objects are represented by the manner in which plates overlap. Nested plates imply a one-to-many relationship between the object classes represented by the outer and inner plates, respectively. Intersecting plates imply a many-to-many relationship between the two object classes represented by the intersecting plates. For example, the plate models in Figure 1a and 1b indicate that objects of type P have a one-to-many relationship with objects of type C. For ease of exposition, in the case of the simple nesting shown in Figure 1a and 1b, we refer to P as a *parent object* and C as a *child object*.

Random variables are represented as circles contained within the plate representing the object class to which the variable corresponds. In Figure 1a and 1b, the variables X and Y characterize objects of type C and the variable Z characterizes objects of type P. *Dependencies* are represented by directed edges between variables. In the context of causal graphical models, edges represent direct causal dependence. Thus, in Figure 1, Z is a common cause of both X and Y.

A *unit* U_i refers to one or more related object classes that will be analyzed simultaneously. A unit instance $u_{i,j}$ is an element of the class U_i. Each row of a typical tabular data set records the values of variables corresponding to a unit instance, and $|U_i| = N$, where N is sample size. This usage follows both statistical and social science practice, where units are sometimes referred to as "units of analysis", "experimental units", or "sampling units". Where clear from context, the term *unit* may also denote a single unit instance $u_{i,j}$.

The simplest unit consists of a single object, such as a person, but units often refer to sets of objects. In general, units are defined as a tuple $\langle \mathcal{O}, \mathcal{R}, \mathcal{P} \rangle$, where $\mathcal{O}$ is a set of object types from which each element of the unit is drawn, $\mathcal{R}$ is the set of relations that hold between object types in $\mathcal{O}$, and $\mathcal{P} \in \mathcal{O}$ is the perspective object type for data analysis (typically, the locus of treatment and outcome). When $|\mathcal{O}| > 1$, we refer to the unit as a *multi-object unit*. For example, a multi-object unit in the domain of Stack Overflow could be defined as $\langle \{author, post\}, \{author \prec post\}, post \rangle$, where $\prec$ indicates a one-to-many relationship. An analysis based on such a unit would study posts, along with characteristics of the authors of those posts. Prior work in relational learning (Maier *et al.*, 2013, 2014) refers to $\mathcal{P}$ as a *perspective*, the term we use here.

Multi-object units are common in statistical and causal analysis, although their existence is often apparent only in the definition of variables. For example, an analysis of academic publishing might appear to use a unit of *paper*, but also include variables such as the impact factor of a paper's *venue* in the analysis, implying a multi-object unit consisting of both papers and venues. The relationship between object classes within a unit can be one-to-one, one-to-many, or many-to-many. For example, venues have a one-to-many relationship with papers, while authors have a many-to-many relationship with papers.

Plate models can represent multi-object units, although $\mathcal{P}$ is typically implicit. The directed structure of typical plate models explicitly defines a generative process for sampling from the joint probability distribution over the values of variables in instances of those multi-object units. Plate models also imply the existence of a generative process over the objects and relationships that constitute the instances of multi-object units, although plate models are typically formalized in ways that assume, rather than define, this process. Koller and Friedman (2009) explicitly define plate models as assuming that a set of objects are given, although it is common for authors to refer to a plate model "generating" a set of object instances.

A single plate model can provide the basis for defining several possible units, because the perspective is only implicit. For example, in the plate model shown in Figure 1a, either the parent P or child C plate could be defined as the perspective for a particular analysis. These could correspond to, respectively, analyses of whether user expertise is a cause of asking more questions ($\mathcal{P}$=user) or whether longer questions cause more answers ($\mathcal{P}$=question).

3.2 DEFINING OBJECT CONDITIONING

Consider the probability space denoted by the tuple $(\Omega, \mathcal{F}, P)$. Ω is a sample space of all possible units[1], $\mathcal{F}$ is a set of events in which each event contains one or more units, and P is a function from events to probabilities. $\mathcal{F}$ is a σ-algebra on Ω, a collection of subsets of units that is closed under complement and under countable unions.

Object conditioning partitions Ω based on an event defined by the identity of one or more object instances contained within those units. Object conditioning on object class O' partitions Ω based on the identity of instances within an object class O' that is a non-perspective member of U' with equal or lower cardinality than the number of instances in the data. Object conditioning thus defines a sub-σ-algebra $\mathcal{F}_{\mathcal{O}'}$. The occurrence of a given object instance o' in a given unit u' corresponds to an event on which a probability distribution or expectation of some other event A can be conditioned, producing $P(A|o' \in u')$ or $E(A|o' \in u')$, respectively.

Conditioning on a discrete variable of O' similarly defines a sub-σ-algebra $\mathcal{F}_{\mathcal{V}}$, in which $\mathcal{F}_{\mathcal{V}} \subseteq \mathcal{F}_{\mathcal{O}}$. That is, object conditioning provides equivalent or finer-grained subsets of Ω than does conditioning on a variable of the same object class. For example, the analysis outlined in Section 2 would partition data records (each of which corresponds to a question) into as many discrete sets as there are *Users*, which is substantially finer-grained than would typically be produced by conditioning on a variable such as the *programming-ability* of a *User*.

Under the assumptions of object conditioning, any variable of an object type on which we condition (e.g., *User*) can take on only a single value within any element of the partition defined by that object type. If that variable is latent, we will not know its value, but the variable can take on only a *single* value for all units in any given subset of the sample space defined by the conditioning event (e.g., questions asked by *user5*).

Object conditioning is useful across a wide range of unit structures. For simplicity of exposition, our focus in this paper is object conditioning in cases in which units $U = \langle\{O_1, O_2\}, \{O_1 \prec O_2\}, O_2\rangle$ consist of objects drawn from two classes O_1 and O_2, O_1 has a one-to-many relationship with O_2, O_2 is the perspective, and we condition on object instances from O_1. In principle, the conditioning event can reference any number of object classes that are present within a unit, but we focus on cases that condition on objects within a single object class. Rattigan (2012) discusses how to apply object conditioning under a wider set of conditions. In addition, all of the effects we discuss here can be demonstrated by plate models with only three variables and three possible dependence structures among those variables—a confounder (common cause), a mediator (causal chain), and a collider (common effect)—even though the effects hold for a much wider class of dependence structures.

Within the framework of causal graphical models, object conditioning is a highly unconventional choice. It is almost never done explicitly. The theoretical infrastructure for *d*-separation and causal identifiability was originally derived only for single-object units. Only recently (Maier *et al.*, 2014; Maier, 2014; Lee and Honavar, 2016) have the underlying principles of *d*-separation been extended to multiple-object units, and this work has shown that the well-known rules of *d*-separation for single-object units do not directly apply to multi-object units. Object conditioning further extends the known dif-

[1]The term *outcome* is typically used, but we use *unit* to avoid confusion with the term used in causal inference to denote a possible effect.

ferences between the analysis of causal dependence in single-object units and multi-object units.

This paper follows an earlier proposal by two of the authors for object conditioning in causal graphical models (Rattigan *et al.*, 2011).[2] However, we now know that the explanation in that paper of the probabilistic implications of object conditioning is both incorrect and incomplete (see Section 5 for details), and the new explanation we propose is one of the principal contributions of this paper.

There exists a special case in which conditioning on objects and conditioning on variables is equivalent. Specifically, conditioning on Z (a variable of the parent object) will approach or be equivalent to conditioning on P (the parent object) as $|Z|$ approaches $|P|$. However, in the vast majority of cases $|Z| \ll |P|$ and $|\mathcal{F}_{\mathcal{V}}| \subset |\mathcal{F}_{\mathcal{O}}|$. Thus the two conditioning operations are only rarely equivalent. In addition, the special case identified above demonstrates another reason to understand the effects of object conditioning. When variable conditioning is equivalent to object conditioning, analysts who are unaware of the effects of object conditioning could observe an apparent set of conditional independencies (X not marginally independent of Y, but X conditionally independent of Y given Z) and incorrectly conclude that Z must be part of a d-connecting path between X and Y. Knowledge of the effects of object conditioning would allow them to correctly infer that the observed evidence only implies that one or more variables on P lie on the d-connecting path between X and Y (and that those variables may not include Z).

3.3 EXCHANGEABILITY

Before discussing the probabilistic implications of object conditioning, we ground the discussion in an unconventional part of probability theory. The traditional discussion of probabilistic implications of specific causal structures (e.g., Pearl, 2009) is grounded in principles of *conditional independence*. Here, we expand the relevant principles to include *conditional exchangeability*.

A sequence of random variables $X_1, \ldots, X_N$ is said to be exchangeable if their joint distribution is invariant under permutations: $F(X_1, \ldots, X_N) = F(X_{\pi(1)}, \ldots, X_{\pi(N)})$ where $\pi(\cdot)$ is a valid permutation of the values $1, \ldots, N$. Informally, the values of X are said to be exchangeable if, given a finite sequence of observations, any re-ordering of this sequence is equally probable. *Conditional* exchangeability is defined analogously to conditional independence: Given a sub-σ-algebra $\mathcal{F}_s$, conditional exchangeability of X is satisfied if, for every element of $\mathcal{F}_s$, X is exchangeable.

The exchangeability of a sequence of random variables is directly relevant to the analysis of observational data. Each unit's version of a random variable X can be thought of as a *separate* (and potentially exchangeable) random variable $X_1, \ldots, X_N$. Note that if these random variables $X_1, \ldots, X_N$ are independent and identically distributed (as often assumed), then this implies exchangeability. However, the converse is not true: exchangeability *does not* imply independence. To understand the difference, consider $X = x_1, x_2, \ldots, x_n$, a finite sequence of independent draws from a Bernoulli distribution with parameter p. If p is known, then the draws are independent. However, if p is unknown, then any set of draws $\{x_1, x_2, \ldots, x_k\}$ where $k < n$ is informative about p and thus provides information about x_{k+1}. The draws are *exchangeable* because any permutation of the indices would leave the distribution of X unaffected, but the draws are not independent if p is unknown (de Finetti, 1931; Kallenberg, 2005; Freer and Roy, 2012; Greenland and Draper, 2011). This latter case is directly analogous to object conditioning, which partitions the space of multi-object units in such a way that Z is known to have only a single value, but the specific value is unknown.

Exchangeability and conditional exchangeability have long been considered a key property for causal inference under the potential outcomes framework (e.g., Greenland and Robins, 1986; Robins and Greenland, 1992). Hernán and Robins (2006) note that "...conditional exchangeability—or some variation of it—is the weakest condition required for causal inference from observational data." The traditional logic of exchangeability in causal inference is relatively straightforward: If treatment X is conditionally exchangeable given Z, then each subset of the sample space defined by some unique value of Z is the equivalent of a randomized experiment. Exchangeability is largely ignored among researchers in causal graphical models in favor of the apparently equivalent property of conditional independence. Below, we show that conditional exchangeability is necessary to understand some cases of object conditioning, and that conditional independence alone cannot explain these cases.

3.4 EXPLAINING OBJECT CONDITIONING

Given the definition of object conditioning, how can we explain its apparent effects? The empirical evidence cited in Section 2 indicates that object conditioning has

[2] The earlier paper used the term *relational blocking*, but we now recommend the term *object conditioning*. The newer term more clearly describes the functioning of the approach. It distinguishes the use of the approach in observational analysis from the use of blocking in experiments. Finally, the term *blocking* has an alternative meaning related to d-separation in graphical models that could prove confusing to some readers.

substantially different probabilistic implications than traditional conditioning on the values of a variable. How can we explain these implications?

We assume the following generative process for multi-object units. Data instances $u_{i,j}$ are drawn randomly from the population of multi-object units. Specifically, for each $u_{i,j}$, a parent object is sampled without replacement from the set of all parent objects and then child objects are randomly sampled from the set of all child objects conditioned on that parent object. Then, variables X, Y, and Z are generated according to a specific causal structure (see below). Each X_i, Y_i, and Z_i is identically distributed, respectively.

Note that the objects and relationships that comprise a unit are generated *prior to* the values of variables. This assumption is implicit in plate models and many other models, in that nearly all such models assume a set of interrelated objects whose variable values are unknown but whose object structure already exists. A violation of this assumption would mean that variables of objects partially or completely determine the relationships among those objects.

Note also that each X_i, Y_i, and Z_i is identically distributed, respectively. Each variable of a child object that is caused by a variable of a parent object is generated with respect to the same value of the variable of the parent object. This assumption is explicit in plate models, but we raise it here to emphasize its importance. Specifically, consider units $U_1 = \langle\{O_1, O_2\}, \{O_1 \prec O_2\}, O_2\rangle$ in which $Z = z_1$ for object $o_{1,j}$ in unit instance $u_{1,j}$ and in which $Z = z_1$ for all instances of O_2 in $u_{1,j}$. That is, the value of Z for an object instance $o_{1,j}$ will remain constant across all child objects related to $o_{1,j}$. A violation of this assumption would mean that different child objects of the same parent object would experience different "versions" of a variable of the parent.

To explore the conditional exchangeability implications of object conditioning, we consider three canonical cases of conditioning that directly match those in Pearl's classic definitions of *d*-separation (Pearl, 2009). Specifically, we examine cases in which the unit is $\langle\{P, C\}, \{P \prec C\}, C\rangle$, in which each unit instance consists of a child object with its associated parent object, and a data set consists of $|C|$ data instances. Within each unit, a variable Z of the parent object P forms the middle node of a three-node structure with two variables X and Y of the child object C. In the three cases, Z corresponds to a *confounder*, a *mediator*, or a *collider*, respectively. Each of these structures is shown in Figure 2.

In each of these cases, we prove that, under object conditioning, the values of X and the values of Y are exchangeable. This directly implies the lack of a biasing path between the treatment X and outcome Y (Flanders and Eldridge, 2015; Angrist *et al.*, 1996; Hernán and Robins, 2006). All three theorems below prove the same property in different contexts.

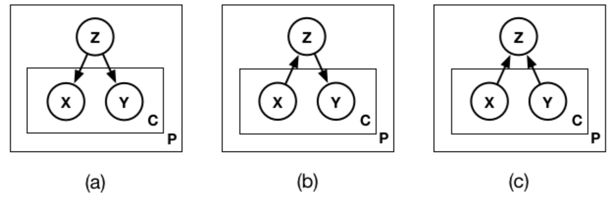

Figure 2: Three simple causal structures in which a variable Z of a parent object serves as a (a) confounder, (b) mediator, or (c) collider, respectively.

Definition 1. *Given the structures of Figure 2, exchangeability among values of X and among values of Y holds if, given two permutations $\pi_x(\cdot)$ and $\pi_y(\cdot)$:*

$$\begin{aligned} &p(x_1, \ldots, x_n, y_1, \ldots, y_n) \\ &\quad = p(x_{\pi_x(1)}, \ldots, x_{\pi_x(n)}, y_1, \ldots, y_n) \\ &\quad = p(x_1, \ldots, x_n, y_{\pi_y(1)}, \ldots, y_{\pi_y(n)}) \\ &\quad = p(x_{\pi_x(1)}, \ldots, x_{\pi_x(n)}, y_{\pi_y(1)}, \ldots, y_{\pi_y(n)}). \end{aligned}$$

Alternatively, $p(X_i, Y_i) = p(X_i, Y_j) = p(X_j, Y_i) = p(X_j, Y_j)$.

First, consider a version of the generative process in which Z is a *confounder* (abbreviated as C1 and illustrated in Figure 2a). Object conditioning on P produces exchangeability among the values of X and among values of Y within the set of child objects of each instance of P. Proofs appear in the supplemental materials.

Theorem 1. *Under C1, object conditioning on P produces exchangeability among values of X and among values of Y.*

That is, no x_i provides special information about the value of the corresponding y_i, thus values of X are exchangeable and values of Y are exchangeable. However, any value x_i provides information about z which, in turn, provides information about y_j, thus X and Y are *not* conditionally independent given the object P, but X and Y *are* conditionally independent given the variable Z.

Next, consider a version of the generative process in which Z is a *mediator* (abbreviated as C2 and illustrated in Figure 2b). Object conditioning on P produces exchangeability among values of X and among values of Y within the set of child objects of each instance of P.

Theorem 2. *Under C2, object conditioning on P produces exchangeability among values of X and among values of Y.*

Again, any value x_i provides information about z which, in turn, provides information about y_j, thus X and Y are *not* conditionally independent given P, but X and Y *are* conditionally independent given Z.

Finally, we consider a version of the generative process in which Z is a *collider* (abbreviated as C3 and illustrated in Figure 2c). Once again, object conditioning on P produces exchangeability among values of X and among values of Y within the set of child objects of each instance of P.

Theorem 3. *Under C3, object conditioning on P produces exchangeability among values of X and among values of Y.*

This exchangeability is a direct consequence of both X and Y being i.i.d. Here, X and Y are marginally independent and remain independent even when conditioned on the object P. In contrast, X and Y are *not* conditionally independent given the variable Z.

3.5 IMPLICATIONS

What does this mean for causal inference? Object conditioning produces exchangeability given all three of the canonical causal structures. These theoretical results are consistent with the empirical evidence cited in Section 2 of Rattigan *et al.* (2011), although the results indicate that the explanation in that earlier paper misinterpreted conditional exchangeability as conditional independence. The proofs also indicate that object conditioning can provide a powerful tool for overcoming latent confounding, one of the most persistent and implacable challenges to effective causal inference. Indeed, where it is applicable, object conditioning can be argued to improve over conventional conditioning on the measured value of a potential confounder on the parent object. First, object conditioning does not run the risk of inducing dependence because of inadvertently conditioning on a collider. Second, it avoids the potential problem of measurement error, which introduces bias into estimates of causal effect (Scheines and Ramsey, 2016).

4 USE OF OBJECT CONDITIONING

We are not the first to recognize the value of object conditioning, although we are the first to accurately explain its surprising probabilistic implications for causal graphical models. As mentioned in Section 1, object conditioning abstractly describes the central features of a family of methods commonly used in social science, economics, medicine, and other fields. Instances from this family include within-subject designs, multi-level models, interrupted time-series designs, and difference-in-differences designs. Each are described in more detail below.

These methods attempt to estimate the strength of causal dependence between treatment and outcome, often denoted as *treatment effect*. We denote the methods below as *designs*, a term used to describe a set of analysis choices that allow stronger causal conclusions than if the design had not been employed. Finally, one of the methods we discuss below—multi-level models—is a modeling formalism rather than a design, but we denote all methods below as *designs* for ease of exposition.

Within-subject designs analyze subjects (e.g., persons) who each receive two or more treatments at different times (Greenwald, 1976). Within-subject designs estimate treatment effect based on a systematic comparison of the outcome values corresponding to the same subject (unit). The example given in Section 2 is a within-subjects design. Closely related design elements include twin studies (Boomsma *et al.*, 2002) and the use of blocking in experimental studies (Fisher, 1935).

Key features of within-subject designs can be represented by object conditioning in which the parent object corresponds to the subject and the child objects correspond to treatment episodes that generate treatment-outcome pairs that vary by episode. Note that some covariates reside on the parent object (e.g., the level of programming knowledge of the writer of a question), rather than on the object corresponding to the treatment episode (e.g., the length of a post).

Multi-level models (MLMs) explicitly analyze units consisting of multiple types of objects nested within one another (Goldstein, 1987, 2010; Gelman and Hill, 2007). They estimate treatment effect by modeling the influence of variables on low-level objects as well as the aggregate effects of higher-level objects, even when the specific extent of higher-level effects is due to latent factors. MLMs are also sometimes referred to as hierarchical models or nested data models.

Key features of multi-level models can be represented as object conditioning, because MLMs typically use only a single variable to correspond to the identity of given higher-level object. The example in Section 2 corresponds directly to a simple MLM.

Interrupted time-series (ITS) designs analyze a temporal sequence of outcome values for each unit over a period in which the value of treatment changes at known times (McDowall *et al.*, 1980). ITS designs estimate treatment effect by comparing, within each unit, the outcome under the new treatment value with the assumed unchanged continuation of the previous outcome trend. Closely related concepts include studies that analyze panel data or longitudinal data, as well as studies that employ cross-over or repeated measures designs.

Key features of ITS designs can be represented as object conditioning, where individual treatment episodes correspond to child objects and the parent object corresponds to an unchanging entity. In Stack Overflow, for example, the parent object could correspond to a user (whose characteristics are assumed to be constant over short time intervals) and the child objects could correspond to short time intervals directly before and after a treatment event such as the awarding of a badge.

Difference-in-differences (DID) designs are similar to ITS designs, but examine outcome values for units that have received multiple values of treatment and units whose value of treatment has not changed (Shadish *et al.*, 2001). They estimate treatment effect by comparing the change in the outcome variable for units experiencing multiple values of treatment with units experiencing only one value. DID designs can be viewed as an elaboration of ITS designs that increases external validity.

Each of the methods above exploits a set of closely related statistical effects that increases the accuracy of (and confidence in) estimates of causal dependence. Each design implicitly conditions its estimate of causal effect for the low-level object on the existence of some higher level object (and the context that object provides for the causal dependence in its associated lower-level objects).

5 PRIOR WORK

Several categories of prior work relate to object conditioning in various ways. Some existing approaches would appear to encompass object conditioning, and we show how each of these approaches fails to fully predict its effects. Other approaches use shared variables to make causal inferences in special cases, and we discuss how these efforts differ substantially from the work discussed here.

The most common approaches to applied causal analysis in multi-object units fall into three general categories: grounding; flattening; and using ID variables. None of these approaches provides a complete and accurate representation of the probabilistic implications provided by object conditioning.

Grounding explicitly uses the semantics of plate models to produce a *ground causal graphical model* given a particular set of object and relationship instances (see the supplemental material for examples of plate models and their corresponding ground models). Indeed, the semantics of plate models are generally tied directly to the ground models resulting from a given combination of plate model and set of object and relationship instances. That is, the meaning of a plate model is as a representation of an infinite set of ground models that it can induce, given a set of related object classes.

However, a ground model explicitly discards the information necessary to express object conditioning in principle or apply the technique in practice. For example, consider the plate model in Figure 1b. Because Z is latent, a ground version of this model would produce only pairs of variables X and Y, but the information about which pairs corresponded to which parent object would be lost. Without that information, we cannot perform object conditioning or obtain its benefits.

Flattening is an alternative to grounding that converts multi-object units into single-object units by importing variables into the perspective object (e.g., Friedman *et al.*, 1999). For example, variables of user objects would be imported into question objects and treated as if they were sampled independently for each question. This converts a plate model of the type in Figure 1b into a plate model of the type in Figure 1c.

As with grounding, flattening fails to represent the opportunities offered by object conditioning. Flattening discards information about the known equivalence among values of Z for child objects belonging to the same parent object, implying instead that each instance of Z for each child is sampled independently. It is possible to condition on the *value* of a variable of a parent object, but flattening loses the information necessary to identify which sets of units necessarily have identical values of Z if it is latent. Thus, flattening cannot express the key properties of object conditioning and must discard all the potential advantages, including eliminating bias due to latent variables of parent objects.

Using ID variables is a final approach which flattens multi-object units, but preserves information about the identity of parent objects by creating a new variable of each child object that records the ID of the parent object. This approach is a core component of multi-level models, which can explicitly estimate parameters for the aggregate effect of individual parent objects on the variables of the corresponding child objects, and it has also been applied as part of various techniques in statistical relational learning (Getoor and Taskar, 2007; Perlich and Provost, 2006). Earlier work by two of the authors (Rattigan *et al.*, 2011; Rattigan, 2012) proposed ID variables as an explanation for the probabilistic implications of object conditioning.

Unlike the prior approaches, ID variables preserve information that enables object conditioning: ID variables can be conditioned on when estimating causal effects. However, treating the identity of the parent object as an ordinary variable assumes that conditioning on such a variable will have the same probabilistic implications as conditioning on other variables. As we show in Section 3, conditioning on objects has different probabilistic implications than conditioning on variables, and thus ID variables will lead to incorrect conclusions. At the very least, ID variables do not share the properties of conventional random variables, and this implies the need for a new syntax and semantics of analytic operations such as conditioning on ID variables.

The explanation in Rattigan *et al.* (2011) attempts to circumvent this objection by assuming that all variables of a parent object deterministically depend on ID, and that the rules of deterministic *d*-separation (Geiger *et al.*, 1990; Spirtes *et al.*, 2000) imply the same conditional independence implications observed in the simulations. However, ID variables lack an internally consistent semantics regarding what can cause ID, what ID can cause, and what intervening on an ID variable means. Furthermore, in the case in which treatment and outcome jointly cause a variable Z of a parent object, it is unclear how an ID variable could (simultaneously) completely determine the value Z (as assumed in that paper). Other researchers (e.g., Theobald, 2015) have agreed with the empirical conclusions of the earlier paper while citing similar doubts about its theoretical explanations. Finally, and quite aside from these problems with ID variables, the explanation in Rattigan *et al.* (2011) contends that conditioning on ID leads to *conditional independence* in cases in which we show that conditional exchangeability is the only property that holds.

Shared variables are the focus of several approaches to causal inference. For example, recent work by Wang and Blei (2018) proposes what they call ”the deconfounder”, an algorithm that infers a latent variable as a substitute for unobserved confounders in multiple-cause settings. This work involves explicit inference of the latent variable, rather than using relational data explicitly for the purposes of conditioning as we do in the work reported here. Balke and Pearl (1994) propose ”twin networks” which posit the existence of background variables that are shared among separate network instances and serve as common causes. These networks were primarily employed as a teaching tool, not a practical tool for analysis. In contrast, we consider the case in which parent objects can participate in a wider array of structures (forks, chains, or colliders) and we propose object conditioning for everyday use in causal inference.

6 CONCLUSIONS

Object conditioning informs a surprisingly diverse set of topics. First, object conditioning provides a new tool to address the persistent challenge of detecting and removing the bias due to latent confounders. If object conditioning substantially reduces or eliminates the observed dependence between treatment and outcome, then (under the assumptions outlined in Section 3) this implies the existence of at least one latent confounder on the parent object. Thus, object conditioning both helps detect the existence of latent confounders on parent objects and provides a means to remove those effects without the attendant risk of inducing collider bias.

Another implication of this work is that some researchers' intuitions about the statistical properties of plate models may be incorrect. For example, when a plate model contains a variable that lies outside of any plate, and the value of that variable is unobserved, the situation is the equivalent of object conditioning. That is, only a single instance of that extra-plate (parent) variable exists, along with many potential instances of the intra-plate (child) variables. In such situations, a latent variable of a parent object that appears to be a confounder will still leave the variables on child objects exchangeable, and yet the naive application of *d*-separation would deem the variables to be marginally dependent.

Finally, this work highlights the utility of conditional exchangeability in causal inference. Object conditioning makes clear that conditional exchangeability can hold in cases that conditional independence does not, and that conditional exchangeability can be an effective tool for inferring which causal models are consistent with the observed data. In prior work on causal inference that did not explicitly consider object conditioning, it could be difficult to identify a clear need for distinguishing between conditional exchangeability and conditional independence. Object conditioning, in contrast, makes the central role of exchangeability in causal inference much more obvious and demonstrates how exchangeability and independence are different yet complementary.

Acknowledgements

The authors thank Marc Maier, Katerina Marazopoulou, David Arbour, Dan Garant, Kaleigh Clary, Amanda Gentzel, and several anonymous UAI reviewers for helpful discussions and comments during the development of this paper. Jensen was supported by the United States Air Force under Contract No. FA8750-17-C-0120. Burroni and Rattigan were supported by the UMass Amherst Center for Data Science. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force or the University of Massachusetts Amherst Center for Data Science.

References

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996.

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, pages 46–54. Morgan Kaufmann Publishers Inc., 1994.

Heejung Bang and James M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Dorret Boomsma, Andreas Busjahn, and Leena Peltonen. Classical twin studies and beyond. *Nature Reviews Genetics*, 3(11):872, 2002.

Wray L. Buntine. Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, 1994.

Diego Colombo, Marloes H. Maathuis, Markus Kalisch, and Thomas S. Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pages 294–321, February 2012.

Bruno de Finetti. *Funzione caratteristica di un fenomeno aleatorio*. Academia Nazionale del Linceo, 1931.

Felix Elwert and Christopher Winship. Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40:31–53, 2014.

R. A. Fisher. *The Design of Experiments*. Oliver & Boyd, 1935.

William D. Flanders and Ronald C. Eldridge. Summary of relationships between exchangeability, biasing paths and bias. *European Journal of Epidemiology*, 30(10):1089–1099, 2015.

Cameron E. Freer and Daniel M. Roy. Computable de Finetti measures. *Annals of Pure and Applied Logic*, 163(5):530–546, 2012.

Nir Friedman, Lise Getoor, Daphne Koller, and Avi Pfeffer. Learning probabilistic relational models. In *IJCAI*, volume 99, pages 1300–1309, 1999.

Dan Geiger, Thomas Verma, and Judea Pearl. Identifying independence in Bayesian networks. *Networks*, 20(5):507–534, 1990.

Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel Hierarchical models*. Cambridge University Press, New York, NY, 2007.

Lise Getoor and Ben Taskar. *Introduction to Statistical Relational Learning*. MIT press, 2007.

Harvey Goldstein. *Multilevel Models in Educational and Social Research*. Griffin, 1987.

Harvey Goldstein. *Multilevel Statistical Models*. Wiley, 4th edition, 2010.

Sander Greenland and David Draper. Exchangeability. In *International Encyclopedia of Statistical Science*, pages 474–476. Springer, 2011.

Sander Greenland and James M. Robins. Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, 15(3):413–419, 1986.

Anthony G. Greenwald. Within-subjects designs: To use or not to use? *Psychological Bulletin*, 83(2):314, 1976.

Miguel A. Hernán and James M. Robins. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*, 60(7):578–586, 2006.

Jennifer L. Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

Guido W. Imbens and Donald B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

Olav Kallenberg. The basic symmetries. *Probabilistic Symmetries and Invariance Principles*, pages 24–68, 2005.

Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT press, 2009.

Sanghack Lee and Vasant Honavar. On learning causal models from relational data. In *AAAI*, pages 3263–3270, 2016.

Marc Maier, Katerina Marazopoulou, David Arbour, and David Jensen. A sound and complete algorithm for learning causal models from relational data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 371–380. AUAI Press, 2013.

Marc Maier, Katerina Marazopoulou, and David Jensen. Reasoning about independence in probabilistic models of relational data. *arXiv preprint arXiv:1302.4381*, 2014.

Marc Maier. *Causal Discovery for Relational Domains: Representation, Reasoning, and Learning*. PhD thesis, University of Massachusetts Amherst, 2014.

Andrew Marder. Stack overflow badges and user behavior: an econometric approach. In *Proceedings of the 12th Working Conference on Mining Software Repositories*, pages 450–453. IEEE Press, 2015.

David McDowall, Richard McCleary, Errol E. Meidinger, and Richard A. Hay Jr. *Interrupted Time Series Analysis.* Sage, 1980.

Hüseyin Oktay, Brian J. Taylor, and David D. Jensen. Causal discovery in social media using quasi-experimental designs. In *Proceedings of the First Workshop on Social Media Analytics*, pages 1–9. ACM, 2010.

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

Judea Pearl. *Causality*. Cambridge University Press, 2009.

Claudia Perlich and Foster Provost. Distribution-based aggregation for relational learning with identifier attributes. *Machine Learning*, 62(1-2):65–105, 2006.

Matthew J. H. Rattigan, Marc Maier, and David Jensen. Relational blocking for causal discovery. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pages 145–151. AAAI Press, 2011.

Matthew J. H. Rattigan. *Leveraging relational representations for causal discovery*. PhD thesis, University of Massachusetts Amherst, 2012.

James M. Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, pages 143–155, 1992.

Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Donald B. Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

Richard Scheines and Joseph Ramsey. Measurement error and causal discovery. In *CEUR workshop proceedings*, volume 1792, page 1. NIH Public Access, 2016.

William R. Shadish, Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth, 2nd edition, 2001.

Amit Sharma, Jake M. Hofman, and Duncan J. Watts. Split-door criterion for causal identification: Automatic search for natural experiments. *arXiv preprint arXiv:1611.09414*, 2016.

Amit Sharma. Necesary and probably sufficient test for finding instrumental variables. *working paper*, 2017.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction and Search*. MIT Press, Cambridge, MA, 2nd edition, 2000.

Roderick Theobald. *Lords Paradox and Targeted Interventions: The Case of Special Education*. PhD thesis, University of Washington, 2015.

Yixin Wang and David M. Blei. The blessings of multiple causes. *arXiv preprint arXiv:1805.06826*, 2018.