Approximate Causal Abstraction

Sander Beckers
Dept. of Philosophy and Religious Studies
Utrecht University
srekcebrednas@gmail.com

Frederick Eberhardt
Humanities and Social Sciences
California Instituite of Technology
fde@caltech.edu

Joseph Y. Halpern
Dept. of Computer Science
Cornell University
halpern@cs.cornell.edu

Abstract

Scientific models describe natural phenomena at different levels of abstraction. Abstract descriptions can provide the basis for interventions on the system and explanation of observed phenomena at a level of granularity that is coarser than the most fundamental account of the system. Beckers and Halpern (2019), building on work of Rubenstein et al. (2017), developed an account of *abstraction* for causal models that is exact. Here we extend this account to the more realistic case where an abstract causal model offers only an approximation of the underlying system. We show how the resulting account handles the discrepancy that can arise between low- and high-level causal models of the same system, and in the process provide an account of how one causal model approximates another, a topic of independent interest. Finally, we extend the account of approximate abstractions to probabilistic causal models, indicating how and where uncertainty can enter into an approximate abstraction.

1 INTRODUCTION

Scientific models aim to provide a description of reality that offers both an explanation of observed phenomena and a basis for intervening on and manipulating the system to bring about desired outcomes. Both of these aims lead to a consideration of models that represent the *causal* relations governing the system. They also imply the need for scientific models that describe the system at a granularity or level of description appropriate for the user and suitable for interventions that are feasible. Such more *abstract* causal models do not capture all the detailed interactions that occur at the most fundamental level of the system, nor do they, in general, represent outcomes of the system completely accurately at the abstract level. Nevertheless, such abstract causal models can (at least) *approximately* explain the phenomena, and can be informative about how the system will respond to interventions that are specified only at the abstract level.

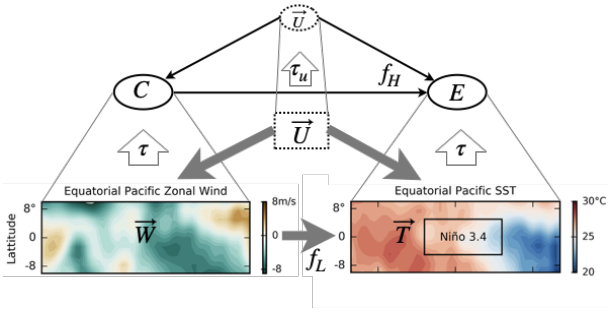

Figure 1: Climate example adapted from Chalupka et al. (2016), in which a high-level causal model for the phenomenon of El Niño is constructed from low-level (high-dimensional) wind $\vec{W}$ and sea surface temperature $\vec{T}$ measurements. $\vec{U}$ is an unmeasured confounder and τ is the mapping between the models. See text for details.

This paper provides a formal account of such approximate abstractions for causal models that builds on the definition of an *abstraction* provided by Beckers and Halpern (2019) (see Section 2), which in turn built on the work of Rubenstein et al. (2017). That notion of abstraction implicitly assumed an underlying causal system that permitted an exact description of the system at the abstract level. Here we weaken that assumption to handle what we take to be the more realistic case, namely, that abstract causal models will capture the underlying system in only an approximate way.

As a simplified working example to illustrate our points we use the case of the wind and sea surface temper-

ature patterns over the equatorial Pacific that give rise to the high-level climate phenomena of El Niño and La Niña, as described by Chalupka et al. (2016). They considered the question of how the El Niño climate phenomenon related to the underlying wind and sea surface temperature patterns that constitute it (see Fig. 1). At the low level they considered two high-dimensional vector-valued variables representing the wind speeds $\vec{W}$ and the sea surface temperatures $\vec{T}$, respectively, on a grid of geographical locations in the equatorial Pacific. They assumed (with some justification from climate science) that wind speed $\vec{W}$ is a cause of sea surface temperature $\vec{T}$, that is, $\vec{T} = f_L(\vec{W}, \vec{U})$ for some high-dimensional function $f_L(.)$ and exogenous causes $\vec{U}$. They allowed the possibility that $\vec{U}$ may be a confounder of $\vec{W}$ and $\vec{T}$, so that there might be an additional causal relation $\vec{W} = g_L(\vec{U})$. Leaving details about feedback and temporal delay aside, they were interested in whether the same system could be described at a higher level, using a low-dimensional structural equation $E = f_H(C, \vec{U})$, where there is a surjective mapping τ from $\mathcal{R}(\vec{T}, \vec{W})$, the set of possible values of $\vec{T}$ and $\vec{W}$, to $\mathcal{R}(E, C)$. In the language of this paper, they were searching for an *abstract* causal description of the system. They required that the high-level model retain a causal interpretation, in the sense that if one intervened on C, there would still be a well-defined causal effect on E, no matter how the intervention on C was interpreted as an intervention on the underlying set of variables $\vec{W}$.

Chalupka et al. (2016) were able to learn such a high-level model, and one of the states of E (the high-level description of the sea surface temperatures) indeed corresponded to what would commonly be described as an El Niño occurring, conventionally defined by an average temperature deviation in a rectangular region of the Pacific. However, the high-level description was not perfect: it provided an informative causal description of the underlying systems and allowed for predictions that *approximated* the actual outcomes. Here we make precise the nature of such an approximation between a high-level and low-level causal model of the same system. In the process, we define what it means for one causal model to approximate another.

Although our running example is a vastly simplified climate model, the challenge of approximately modeling phenomena at a more abstract level is part of almost *every* scientific model. For example, it was Robert Boyle's great insight that, despite its inaccuracies for real gases in practice, the ideal gas law still provides an *approximate* abstract description of the behavior of the molecules of a gas in a container that is extraordinarily useful for understanding and manipulating real systems. Approximate abstractions can take a variety of forms in scientific practice, ranging from idealizations and discretizations to other forms of simplification and dimension reduction (as in the climate example). Our account captures these in a unified formal framework.

The main contribution of this paper is to present a framework that offers a foundation for analyzing abstraction and approximation in causal models. We provide what we believe are sensible definitions of approximation and approximate abstraction, and a conceptual discussion of these notions. In addition, we provide some technical results regarding the difficulty of determining whether an approximate abstraction can be viewed as the composition of an approximation and an exact abstraction.

2 PRELIMINARIES

Since we are interested in scientific models that support explanations of phenomena and can inform interventions on a system, we start by defining a deterministic causal model with a set of possible interventions. We use exogenous and endogenous variables to distinguish those influences that are external to the system and those that are internal. The definitions follow the framework developed by Halpern (2016).

Definition 2.1: A signature $\mathcal{S}$ is a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{R})$, where $\mathcal{U}$ is a set of *exogenous* variables, $\mathcal{V}$ is a set of *endogenous* variables, and $\mathcal{R}$, a function that associates with every variable $Y \in \mathcal{U} \cup \mathcal{V}$ a nonempty set $\mathcal{R}(Y)$ of possible values for Y (i.e., the set of values over which Y *ranges*). If $\vec{X} = (X_1, \ldots, X_n)$, $\mathcal{R}(\vec{X})$ denotes the crossproduct $\mathcal{R}(X_1) \times \cdots \times \mathcal{R}(X_n)$. ∎

For simplicity in this paper, we assume that signatures are finite, that is, $\mathcal{U}$ and $\mathcal{V}$ are finite, and the range of each variable $Y \in \mathcal{U} \cup \mathcal{V}$ is finite.

Definition 2.2: A *basic causal model* M is a pair $(\mathcal{S}, \mathcal{F})$, where $\mathcal{S}$ is a signature and $\mathcal{F}$ defines a function that associates with each endogenous variable X a *structural equation* F_X giving the value of X in terms of the values of other endogenous and exogenous variables. Formally, the equation F_X maps $\mathcal{R}(\mathcal{U} \cup \mathcal{V} - \{X\})$ to $\mathcal{R}(X)$, so F_X determines the value of X, given the values of all the other variables in $\mathcal{U} \cup \mathcal{V}$. ∎

Note that there are no functions associated with exogenous variables, since their values are determined outside the model. We call a setting $\vec{u}$ of values of exogenous variables a *context*.[1]

[1] We remark that the notion of context used here, which goes back to (Halpern and Pearl 2005), is similar to that of Boutilier

The value of X may not depend on the values of all other variables. Y *depends on* X *in context* $\vec{u}$ if there is some setting of the endogenous variables other than X and Y such that if the exogenous variables have value $\vec{u}$, then varying the value of X in that context results in a variation in the value of Y; that is, there is a setting $\vec{z}$ of the endogenous variables other than X and Y and values x and x' of X such that $F_Y(x, \vec{z}, \vec{u}) \neq F_Y(x', \vec{z}, \vec{u})$.

In this paper, we restrict attention to *recursive* (or *acyclic*) models, that is, models where there is a partial order $\preceq$ on variables such that if Y depends on X, then $X \prec Y$.[2] In a recursive model, given a context $\vec{u}$, the values of all the remaining variables are determined (we can just solve for the value of the endogenous variables in the order given by $\prec$) We often write the equation for an endogenous variable as $X = f(\vec{Y})$; this denotes that the value of X depends only on the values of the variables in $\vec{Y}$, and the connection is given by f. Our climate example is recursive, since $\vec{T} = f_L(\vec{W}, \vec{U})$.

An *intervention* has the form $\vec{X} \leftarrow \vec{x}$, where $\vec{X}$ is a set of endogenous variables. Intuitively, this means that the values of the variables in $\vec{X}$ are set to $\vec{x}$. The structural equations define what happens in the presence of interventions. Setting the value of some variables $\vec{X}$ to $\vec{x}$ in a causal model $M = (\mathcal{S}, \mathcal{F})$ results in a new causal model, denoted $M_{\vec{X} \leftarrow \vec{x}}$, which is identical to M, except that $\mathcal{F}$ is replaced by $\mathcal{F}^{\vec{X} \leftarrow \vec{x}}$: for each variable $Y \notin \vec{X}$, $F_Y^{\vec{X} \leftarrow \vec{x}} = F_Y$, while for each X' in $\vec{X}$, the equation $F_{X'}$ for is replaced by $X' = x'$ (where x' is the value in $\vec{x}$ corresponding to $\vec{x}$).

Halpern and Pearl (2005) and Halpern (2016) implicitly assumed that all interventions can be performed in a model. For reasons that will become clear when defining abstraction, we follow Rubenstein et al. (2017) and Beckers and Halpern (2019) in adding the notion of "allowed interventions" to a causal model. This allows us to capture situations where not all interventions are of interest to the modeler and/or some interventions may not be feasible. We can then define a *causal model* M as a tuple $(\mathcal{S}, \mathcal{F}, \mathcal{I})$, where $(\mathcal{S}, \mathcal{F})$ is a basic causal model and $\mathcal{I}$ is a set of *allowed interventions*. We sometimes write a causal model $M = (\mathcal{S}, \mathcal{F}, \mathcal{I})$ as $(M', \mathcal{I})$, where M' is the basic causal model $(\mathcal{S}, \mathcal{F})$, if we want to emphasize the role of the allowed interventions.

Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, a *primitive event* is a formula of the form $X = x$, for $X \in \mathcal{V}$ and $x \in \mathcal{R}(X)$. A *causal formula (over* $\mathcal{S}$*)* is one of the form $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi$, where φ is a Boolean combination of primitive events, $Y_1, \ldots, Y_k$ are distinct variables in $\mathcal{V}$, and $y_i \in \mathcal{R}(Y_i)$. Such a formula is abbreviated as $[\vec{Y} \leftarrow \vec{y}]\varphi$. The special case where $k = 0$ is abbreviated as φ. Intuitively, $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi$ says that φ would hold if Y_i were set to y_i, for $i = 1, \ldots, k$.

A causal formula ψ is true or false in a causal model, given a context. As usual, we write $(M, \vec{u}) \models \psi$ if the causal formula ψ is true in causal model M given context $\vec{u}$. The $\models$ relation is defined inductively. $(M, \vec{u}) \models X = x$ if the variable X has value x in the unique (since we are dealing with recursive models) solution to the equations in M in context $\vec{u}$ (i.e., the unique vector of values that simultaneously satisfies all equations in M with the variables in $\mathcal{U}$ set to $\vec{u}$). The truth of conjunctions and negations is defined in the standard way. Finally, $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ if $(M_{\vec{Y} = \vec{y}}, \vec{u}) \models \varphi$.

To simplify notation, we sometimes write $M(\vec{u})$ to denote the unique element of $\mathcal{R}(\mathcal{V})$ such that $(M, \vec{u}) \models \mathcal{V} = \vec{v}$. Similarly, given an intervention $\vec{Y} \leftarrow \vec{y}$, $M(\vec{u}, \vec{Y} \leftarrow \vec{y})$ denotes the unique element of $\mathcal{R}(\mathcal{V})$ such that $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}](\mathcal{V} = \vec{v})$.

These definitions allow us to describe the climate system both in terms of a low-level causal model M_L and a high-level model M_H:

$$\begin{aligned} M_L &= ((\mathcal{U}_L, \mathcal{V}_L, \mathcal{R}_L), \mathcal{F}_L, \mathcal{I}_L) \\ &= ((\{\vec{U}\}, \{\vec{W}, \vec{T}\}, \mathcal{R}_L), \{g_L, f_L\}, \mathcal{I}_L) \text{ and} \\ M_H &= ((\mathcal{U}_H, \mathcal{V}_H, \mathcal{R}_H), \mathcal{F}_H, \mathcal{I}_H) \\ &= ((\{\vec{U}\}, \{C, E\}, \mathcal{R}_H), \{g_H, f_H\}, \mathcal{I}_H). \end{aligned}$$

Chalupka et al. (2016) treated $\vec{U}$ as not only exogenous, but also unobserved, and therefore made no claim about its dimensionality in the high-level model. We do not explicitly spell out $\mathcal{R}_L$ and $\mathcal{R}_H$ here; the description of the climate model indicates that $\mathcal{R}_L(C) \times \mathcal{R}_L(E)$ is much smaller than $\mathcal{R}_L(\vec{W}) \times \mathcal{R}_L(\vec{T})$: many different low-level (vector-valued) states may correspond to one high-level low-dimensional state. Finally, for our climate example we have not yet said anything about interventions, so $\mathcal{I}_L$ and $\mathcal{I}_H$ are currently placeholders.

2.1 Abstraction

Beckers and Halpern (2019) gave a sequence of successively more restrictive definitions of abstraction for causal model. The first and least restrictive definition is the notion of *exact transformation* due to Rubenstein et al. (2017). Examples given by Beckers and Halpern show that the notion of exact transformation is arguably

et al. (1996), in that both are assignments of values to variables. However, here it is used in particular to denote an assignment of values to all the exogenous variables.

[2] Halpern (2016) calls this *strongly recursive*, in order to distinguish it from models in which the partial order depends on the context. This distinction has no impact on our results.

too flexible. Thus, in this paper the notion of abstraction we consider is that of *τ-abstractions*, introduced by Beckers and Halpern, which can be viewed as a restriction of exact transformations[3] and avoids some of their problems. However, nothing hinges on this choice: all definitions can just as well be interpreted using the less restrictive notions of abstractions.

The key to defining all the notions of abstraction from a low-level to a high-level causal model considered by Beckers and Halpern is the abstraction function $\tau : \mathcal{R}_L(\mathcal{V}_L) \to \mathcal{R}_H(\mathcal{V}_H)$ that maps endogenous states of M_L to endogenous states of M_H. This is a generalization of the surjective mapping τ from $\mathcal{R}(\vec{T}, \vec{W})$ to $\mathcal{R}(E, C)$ discussed in the introduction. In the formal definition, we need two additional functions: $\tau_{\mathcal{U}} : \mathcal{R}_L(\mathcal{U}_L) \to \mathcal{R}_H(\mathcal{U}_H)$, which maps exogenous states of M_L to exogenous states of M_H, and $\omega_\tau : \mathcal{I}_L \to \mathcal{I}_H$, which maps low-level interventions to high-level interventions. Beckers and Halpern (2019) show that, given their definition of abstraction, $\tau_{\mathcal{U}}$ and ω_τ can be derived from τ. We briefly review the relevant definitions here; we refer the reader to their paper for more details and motivation.

Definition 2.3: Given a set $\mathcal{V}$ of endogenous variables, $\vec{X} \subseteq \mathcal{V}$, and $\vec{x} \in \mathcal{R}(\vec{X})$, let

$$Rst(\mathcal{V}, \vec{x}) = \{\vec{v} \in \mathcal{R}(\mathcal{V}) : \vec{x} \text{ is the restriction of } \vec{v} \text{ to } \vec{X}\}.$$

Given $\tau : \mathcal{R}_L(\mathcal{V}_L) \to \mathcal{R}_H(\mathcal{V}_H)$, define $\omega_\tau(\vec{X} \leftarrow \vec{x}) = \vec{Y} \leftarrow \vec{y}$ if $\vec{Y} \subseteq \mathcal{V}_H$, $\vec{y} \in \mathcal{R}_H(\vec{Y})$, and $\tau(Rst(\mathcal{V}_L, \vec{x})) = Rst(\mathcal{V}_H, \vec{y})$ (where, as usual, given $T \subseteq \mathcal{R}_L(\mathcal{V}_L)$, define $\tau(T) = \{\tau(\vec{v}_L) : \vec{v}_L \in T\}$). It is easy to see that, given $\vec{X}$ and $\vec{x}$, there can be at most one such $\vec{Y}$ and $\vec{y}$. If such a $\vec{Y}$ and $\vec{y}$ do not exist, we take $\omega_\tau(\vec{X} \leftarrow \vec{x})$ to be undefined. Let $\mathcal{I}_L^\tau$ be the set of interventions for which ω_τ is defined, and let $\mathcal{I}_H^\tau = \omega_\tau(\mathcal{I}_L^\tau)$. ∎

Note that if τ is surjective, then it easily follows that $\omega_\tau(\emptyset) = \emptyset$, and for all $\vec{v}_L \in \mathcal{R}_L(\mathcal{V}_L)$, $\omega_\tau(\mathcal{V}_L \leftarrow \vec{v}_L) = \mathcal{V}_H \leftarrow \tau(\vec{v}_L)$.

With this definition, the need for the intervention sets $\mathcal{I}_L$ and $\mathcal{I}_H$ becomes clear: in general, not all low-level interventions will neatly map to a high-level intervention, since the abstraction function τ may aggregate variables together; some low-level interventions will constitute only a partial intervention on a high-level variable. The "allowed intervention" sets ensure that the set of interventions can be suitably restricted to retain only those that can actually be abstracted. Similarly, there may be cases where the high-level model does not support all interventions because they may not be well-defined. For example, what does it mean in the ideal gas law to change temperature, while keeping pressure and volume constant? It is not even clear that such an intervention is meaningful.

Of course, a minimal requirement for any causal model to be a τ-abstraction of some other model is that the signatures of both models need to be compatible with τ. Beckers and Halpern (2019) add two further minimal requirements. We capture all of them by requiring the two causal models to be τ-consistent:

Definition 2.4 : If $\tau : \mathcal{R}_L(\mathcal{V}_L) \to \mathcal{R}_H(\mathcal{V}_H)$, then $(M_H, \mathcal{I}_H)$ and $(M_L, \mathcal{I}_L)$ are *τ-consistent* if τ is surjective, $\mathcal{I}_H = \omega_\tau(\mathcal{I}_L)$, and $|\mathcal{R}_L(\mathcal{U}_L)| \geq |\mathcal{R}_H(\mathcal{U}_H)|$. ∎

Definition 2.5 : $(M_H, \mathcal{I}_H)$ is a *τ-abstraction* of $(M_L, \mathcal{I}_L)$ if $(M_H, \mathcal{I}_H)$ and $(M_L, \mathcal{I}_L)$ are τ-consistent and there exists a surjective $\tau_{\mathcal{U}}$ such that for all $\vec{u}_L \in \mathcal{R}_L(\mathcal{U}_L)$ and $\vec{X} \leftarrow \vec{x} \in \mathcal{I}_L$, $\tau(M_L(\vec{u}_L, \vec{X} \leftarrow \vec{x})) = M_H(\tau_{\mathcal{U}}(\vec{u}_L), \omega_\tau(\vec{X} \leftarrow \vec{x}))$. ∎

Abstraction means that for each possible low-level context-intervention pair, the two ways of moving up "diagonally" to a high-level endogenous state always lead to the same result. The first way is to start by applying M_L to get a low-level state, and then moving up to a high-level state by applying τ, whereas the second way is to first move to a high-level context and intervention (by applying $\tau_{\mathcal{U}}$ and ω_τ), and then to obtain a high-level state by applying M_H.

A common and useful form of abstraction occurs when the low-level variables are clustered, so that the clusters form the high-level variables. Roughly speaking, the intuition is that in the high-level model, one variable captures the effect of a number of variables in the low-level model. This makes sense only if the low-level variables that are being clustered together "work the same way" as far as the allowed interventions go. The following definition makes this special case of abstraction precise.

Definition 2.6 : If $\mathcal{V}_H = \{Y_1, \ldots, Y_n\}$, then $\tau : \mathcal{R}_L(\mathcal{V}_L) \to \mathcal{R}_H(\mathcal{V}_H)$ is *constructive* if there exists a partition $P = \{\vec{Z}_1, \ldots, \vec{Z}_{n+1}\}$ of $\mathcal{V}_L$, where $\vec{Z}_1, \ldots, \vec{Z}_n$ are nonempty, and mappings $\tau_i : \mathcal{R}(\vec{Z}_i) \to \mathcal{R}(Y_i)$ for $i = 1, \ldots, n$ such that $\tau = (\tau_1, \ldots, \tau_n)$; that is, $\tau(\vec{v}_L) = \tau_1(\vec{z}_1) \cdot \ldots \cdot \tau_n(\vec{z}_n)$, where $\vec{z}_i$ is the projection of $\vec{v}_L$ onto the variables in $\vec{Z}_i$, and $\cdot$ is the concatenation operator on sequences. If M_H is a τ-abstraction of

[3] Exact τ-transformations relate probabilistic causal models, while τ-abstractions relate (deterministic) causal models. Beckers and Halpern (2019) show that we can compare the two by proving that every τ-abstraction is what they call a *uniform τ-transformation*: specifically, if M_H is a τ-abstraction of M_L, then for every probability $\Pr_L$, there exists a probability $\Pr_H$ such that $(M_L, \Pr_L)$ is an exact τ-transformation of $(M_H, \Pr_H)$.

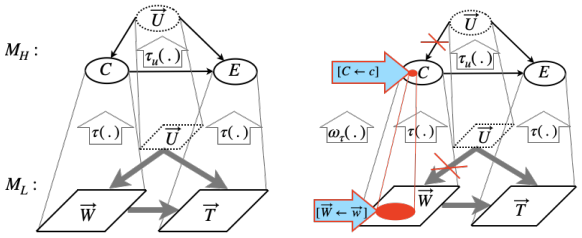

Figure 2: The climate model exemplifies constructive abstraction. *Left:* τ and $\tau_{\mathcal{U}}$ map low-level variables to their high-level counterparts. *Right:* ω_τ maps low-level interventions to the high-level intervention. Several different low-level interventions on $\vec{W}$ may correspond to the same intervention on C.

M_L then we say it is *constructive* if τ is constructive and $\mathcal{I}_L = \mathcal{I}_L^\tau$. ∎

In this definition, we can think of each $\vec{Z}_i$ as describing a set of microvariables that are mapped to a single macrovariable Y_i. The variables in $\vec{Z}_{n+1}$ (which might be empty) are ones that are marginalized away.

The climate example almost exactly fits the notion of constructive abstraction: the variables in the high-level model, C and E, each correspond to a vector-valued low-level variable, $\vec{W}$ and $\vec{T}$, but $\vec{W}$ and $\vec{T}$ could have been replaced by disjoint sets of variables. Consequently, τ maps states from the same low-level variable to the same high-level variable (see Fig. 2, left). Although interventions are practically not feasible in the climate case, hypothetically they are perfectly well-defined: an intervention on C can be instantiated at the low level by several different interventions on $\vec{W}$ (see Fig. 2, right). Finally, given an intervention on $\vec{W}$, we have that

$$M_L(\vec{u}, \vec{W} \leftarrow \vec{w}) \models (\vec{T} = \vec{t}) \text{ iff}$$
$$M_H(\tau_{\mathcal{U}}(\vec{u}), \omega_\tau(\vec{W} \leftarrow \vec{w})) \models (E = \tau(\vec{t})).$$

The correspondence between M_L and M_H is *exact*. In fact, the high-level model M_H constructed by Chalupka et al. (2016) did not satisfy this biconditional precisely, but had to approximate it. We maintain that, in general, high-level models in science are only *approximate abstractions*.

3 APPROXIMATE ABSTRACTION

In order to define what it means for one causal model to be an approximation of another, we need a way of measuring the "distance" between causal models. We take a *distance function* to simply be a function that associates with a pair (M_1, M_2) a distance, that is, a non-negative real number. We show how various distance functions on causal models can be defined, starting from a *metric* $d_{\mathcal{V}}$ on the state space $\mathcal{R}(\mathcal{V})$ of a causal model. (Recall that a metric on a space X is a function $d : X \times X \to \mathbb{R}^+$ such that (a) $d(x, x') = 0$ iff $x = x'$, (b) $d(x, y) = d(y, x)$, and (c) $d(x, y) + d(y, z) \geq d(x, z)$.) Such a metric $d_{\mathcal{V}}$ is typically straightforward to define. Given two states s_1 and s_2, we can compare the value of each endogenous variable X in s_1 and s_2. The difference in the values determines the distance between s_1 and s_2.

The choice of distance function is application-dependent. Different researchers looking at the same data may be interested in different aspects of the data. For example, suppose that the model is defined in terms of 5 variables, $X_1, \ldots, X_5$. X_3 might be gender and X_4 might be height. Suppose that we restrict to distance functions that takes the distance between $(x_1, \ldots, x_5)$ and $(x'_1, ..., x'_5)$ to be of the form $\sqrt{w_1(x_1 - x'_1)^2 + \cdots + w_5(x_5 - x'_5)^2}$, where $w_1, \ldots, w_5$ are weights that represent the importance (to the researcher) of each of these five features. One researcher might not be interested in gender (so doesn't care if her predictions about gender are incorrect), and thus might take $w_3 = 0$; another researcher might care about gender and not about height, so she might take $w_3 = 1$ and $w_4 = 0$. While, as we shall see, the choice of distance function makes a crucial difference in evaluating the "goodness" of an approximate abstraction, in light of the above, we leave the choice of distance function unspecified.

In the remainder of the paper, we assume that the state space $\mathcal{R}(\mathcal{V})$ of endogenous variables for each causal model comes with a metric $d_{\mathcal{V}}$. We provide a number of ways of lifting the metric $d_{\mathcal{V}}$ on states to a distance function d on models, and then use the distance function to define both approximation and approximate abstraction.

Our intuition for the distance function is based on how causal models are typically used. Specifically, we are interested in how two models compare with regard to the predictions they make about the effects of an intervention. Our intuition is similar in spirit to that behind the notion of *structural intervention distance* considered by Peters and Bühlmann (2015), although the technical definitions are quite different. (We discuss the exact relationship between our approach and theirs in the next section, in the context of probabilistic models.)

We start with the simplest setting where this intuition can be made precise, one where the models M_1 and M_2 differ only with regards to their equations. We say that two models are *similar* in this case. If two models are similar, then, among other things, we can assume that they have the same metric $d_{\mathcal{V}}$. In this setting, we can compare the effect of each allowed intervention $\vec{X} \leftarrow \vec{x}$ in the two

models. That is, for each context $\vec{u}$, we can compare the states $M_1(\vec{u}, \vec{X} \leftarrow \vec{x})$ and $M_2(\vec{u}, \vec{X} \leftarrow \vec{x})$ that arise after performing the intervention $\vec{X} \leftarrow \vec{x}$ in context $\vec{u}$ in each model. We get the desired distance function by taking the worst-case distance between all such states.

Definition 3.1: Define a distance function d_{max} on pairs of similar models by taking

$$d_{max}(M_1, M_2) = \max_{\vec{X} \leftarrow \vec{x} \in \mathcal{I},\, \vec{u} \in \mathcal{R}(\mathcal{U})} \{d_{\mathcal{V}}(M_1(\vec{u}, \vec{X} \leftarrow \vec{x}), M_2(\vec{u}, \vec{X} \leftarrow \vec{x}))\}.$$

The causal model M_1 is a d_{max}-α *approximation* of M_2 if $d_{max}(M_1, M_2) \leq \alpha$. ∎

Thus, M_1 is a d_{max}-α approximation of M_2 if the predictions of M_1 are always within α of the predictions of M_2.

We apply similar ideas to defining approximate abstraction. But now we no longer have a distance function defined on causal models with the same signature. Rather, the distance function d_τ is defined on pairs (M_L, M_H) consisting of a low-level and high-level causal model (which, in general, have different signatures), related by a a surjective mapping τ. The idea behind d_τ is that we start with a low-level intervention $\vec{X} \leftarrow \vec{x}$, consider its effects in M_L, lift this up to M_H using τ, and compare this to the effects of $\omega_\tau(\vec{X} \leftarrow \vec{x})$ in M_H.

Definition 3.2: Fix a surjective map $\tau : \mathcal{R}(\mathcal{V}_L) \rightarrow \mathcal{R}(\mathcal{V}_H)$. Define the distance function d_τ on pairs of τ-consistent models (M_L, M_H) by taking

$$d_\tau(M_L, M_H) = \min_{\tau_{\mathcal{U}} \text{ surjective}} \max_{\vec{X} \leftarrow \vec{x} \in \mathcal{I}_L,\, \vec{u}_L \in \mathcal{R}_L(\mathcal{U}_L)} (d_{\mathcal{V}_H}(\tau(M_L(\vec{u}_L, \vec{X} \leftarrow \vec{x})), M_H(\tau_{\mathcal{U}}(\vec{u}_L), \omega_\tau(\vec{X} \leftarrow \vec{x})))).$$

M_H is a τ-α *approximate abstraction* of M_L if $d_\tau(M_L, M_H) \leq \alpha$. ∎

We take the minimum over all functions $\tau_{\mathcal{U}}$ because the function that lifts the low-level contexts up to the high-level contexts does not play a major role. We thus simply focus on the best choice of $\tau_{\mathcal{U}}$.

To get an intuition for an approximate abstraction, consider the climate example again. For a low-level intervention $\vec{W} \leftarrow \vec{w}$ and a low-level context $\vec{u}$, there are two ways of lifting their effect to M_H (see Fig. 3). The first is to start by applying M_L to the context-intervention pair to determine a low-level state $\vec{t}$, and then apply τ to obtain the high-level state $E = e = \tau(\vec{t})$. (Recall that M_L can be viewed as a function from context-intervention pairs to states in $\mathcal{R}_L(\mathcal{V}_L)$.) The second is to first lift the intervention $\vec{W} \leftarrow \vec{w}$ to $\mathcal{I}_H$, that is, an intervention on C, and the context to $\mathcal{U}_H$ by applying ω_τ and $\tau_{\mathcal{U}}$.

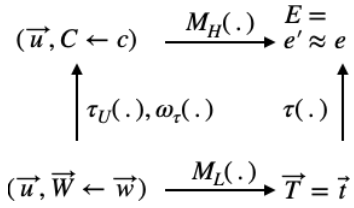

Figure 3: Approximate abstraction in the climate example. We measure (for the worst-case choice of low-level intervention $\vec{W} \leftarrow \vec{w}$ and context $\vec{u}$) the distance between the values of the variable E obtained by (a) first applying M_L to $\vec{u}$ and $\vec{W} \leftarrow \vec{w}$ and then abstracting by applying τ vs. (b) abstracting the intervention and context (by applying ω_τ and $\tau_{\mathcal{U}}$, respectively) and then applying M_H.

Then we apply $M_H(\vec{u}_H, C \leftarrow c)$, which again gives a high-level endogenous state e'. We are identifying the degree to which M_H approximates M_L with the worst-case distance between the two ways of lifting the context-intervention pairs, for an optimal choice of $\tau_{\mathcal{U}}$.

The following straightforward results show that our notion of approximate abstraction is a sensible generalization of both the notion of an exact abstraction and the notion of approximation between similar models.

Proposition 3.3: *M_H is a τ-0-approximate abstraction of M_L iff M_H is a τ-abstraction of M_L.*

Proposition 3.4: *If M_1 and M_2 are similar, then M_2 is a Id-α-approximate abstraction of M_1 (where Id is the identity function on $\mathcal{R}(\mathcal{V})$) iff M_2 is a d_{max}-α-approximation of M_1.*

4 APPROXIMATE ABSTRACTION FOR PROBABILISTIC CAUSAL MODELS

A *probabilistic causal model* $M = (\mathcal{S}, \mathcal{F}, \mathcal{I}, \Pr)$ is just a causal model together with a probability $\Pr$ on contexts $\vec{u}$. In this section, we assume that all causal models are probabilistic, and extend the notion of approximation to probabilistic causal models. We again start by considering the simplest setting, where we have probabilistic models that differ only in their equations. We again call such models *similar*. Now we have several reasonable distance functions.

Definition 4.1: Define a distance function d_{max} on pairs of similar probabilistic causal models by taking

$$d_{max}(M_1, M_2) = \max_{\vec{X} \leftarrow \vec{x} \in \mathcal{I}} \left(\sum_{\vec{u} \in \mathcal{R}(\mathcal{U})} \Pr(\vec{u}) d_{\mathcal{V}}(M_1(\vec{u}, \vec{X} \leftarrow \vec{x}), M_2(\vec{u}, \vec{X} \leftarrow \vec{x})) \right).$$

The probabilistic causal model M_1 is a d_{max}-α *approximation* of M_2 if $d_{max}(M_1, M_2) \leq \alpha$. ∎

Here we have just replaced the max over contexts in Definition 3.1 by an expectation over contexts.

But we may not always just be interested in the expected distance. We may, for example, be more concerned with the likelihood of serious prediction differences, and not be too concerned about small differences. This leads to the following definition.

Definition 4.2: Define a distance function d_β on pairs of similar probabilistic causal models by taking

$$d_\beta(M_1, M_2) = \max_{\vec{X} \leftarrow \vec{x} \in \mathcal{I}} \Pr(\{\vec{u} : d_{\mathcal{V}}(M_1(\vec{u}, \vec{X} \leftarrow \vec{x}), M_2(\vec{u}, \vec{X} \leftarrow \vec{x})) \geq \beta\}.$$

The probabilistic causal model M_1 is a d_β-α *approximation* of M_2 if $d_\beta(M_1, M_2) \leq \alpha$. ∎

We can now extend these ideas to approximate abstraction. We first extend the definition of τ-abstraction to the probabilistic setting.

Note that we can view Pr as a probability measure on $\mathcal{R}(\mathcal{V})$, by taking $\Pr(\vec{v}) = \{\vec{u} : M(\vec{u}) = \vec{v}\}$. An intervention $\vec{X} \leftarrow \vec{x}$ also induces a probability $\Pr^{\vec{X} \leftarrow \vec{x}}$ on $\mathcal{R}(\mathcal{V})$ in the obvious way:

$$\Pr^{\vec{X} \leftarrow \vec{x}}(\vec{v}) = \Pr(\{\vec{u} : M(\vec{u}, \vec{X} \leftarrow \vec{x}) = \vec{v}\}).$$

In the deterministic notion of abstraction, we require that the two high-level states obtained by the two different ways of lifting the effects of a low-level intervention to the high level be equal. In the probabilistic notion, we require that the two different probability distributions obtained by the two ways of lifting an intervention be equal.

Definition 4.3: M_H is a τ-*abstraction* of M_L if M_H and M_L are τ-consistent and for all interventions $\vec{X} \leftarrow \vec{x} \in \mathcal{I}_L$, we have that $\tau(\Pr_L^{\vec{X} \leftarrow \vec{x}}) = \Pr_H^{\omega_\tau(\vec{X} \leftarrow \vec{x})}$. ∎

We can now extend our definitions to the approximate scenario just as we did for deterministic causal models.

Definition 4.4: Fix a surjective map $\tau : \mathcal{R}(\mathcal{V}_L) \rightarrow \mathcal{R}(\mathcal{V}_H)$. Define the distance function d_τ on pairs of τ-consistent probabilistic causal models by taking

$$d_\tau(M_L, M_H) = \min_{\{\tau_{\mathcal{U}} : \tau_{\mathcal{U}}(\Pr_L) = \Pr_H\}} \max_{\vec{X} \leftarrow \vec{x} \in \mathcal{I}_L} (\sum_{\vec{u}_L \in \mathcal{R}_L(\mathcal{U}_L)} \Pr_L(\vec{u}_L) d_{\mathcal{V}_H}(\tau(M_L(\vec{u}_L, \vec{X} \leftarrow \vec{x})), M_H(\tau_{\mathcal{U}}(\vec{u}_L), \omega_\tau(\vec{X} \leftarrow \vec{x})))).$$

M_H is a τ-α *approximate abstraction* of M_L if $d_\tau(M_L, M_H) \leq \alpha$. ∎

For the climate example this definition implies the following: Suppose we introduce probabilities by specifying distributions over the contexts $\vec{u}_L$ and $\vec{u}_H$. The resulting probabilistic causal model (M_H, Pr_H) is a τ-α approximate abstraction of (M_L, Pr_L) if the expectation (in terms of Pr_L) of the difference between the states e and e' of the high-level temperature variable E is less or equal to α, where e and e' are determined exactly in accordance with the two pathways in Fig. 3, selecting the worst-case intervention $\vec{W} \leftarrow \vec{w}$ and the best case $\tau_{\mathcal{U}}$.

Analogously to the deterministic case, we have the following straightforward results.

Proposition 4.5: *M_H is a τ-0-approximate abstraction of M_L iff M_H is a τ-abstraction of M_L.*

Proposition 4.6: *If M_1 and M_2 are similar, then M_2 is a Id-α-approximate abstraction of M_1 iff M_2 is a d_{max}-α-approximation of M_1.*

In Definition 4.4 we consider the worst-case low-level intervention to define the distance. In many cases, however, the whole point of an abstraction is to be able to exclude rare low-level boundary cases (e.g., when the ideal gas law is taken to refer only to equilibrium states). Moreover, often the actual manipulations that we can perform are known to us only at the high-level, because the low-level implementation of the intervention is unobservable to us. For example, in setting the room temperature to 70°F we do not generally consider the instantiation of that intervention which superheats one corner of the room and freezes the rest such that the mean kinetic energy works out just right. As Spirtes and Scheines (2004) show, in the absence of any further information, such ambiguous manipulations can be quite problematic. Fortunately, often our knowledge of the mechanism of how a high-level intervention is implemented does give us significant probabilistic information. For example, we might know that the heater has a fan that circulates the air, most likely resulting in relatively uniform distributions of the kinetic energies of the particles.

We capture this information using what we call an intervention distribution.

Definition 4.7 : Given a surjective map $\tau : \mathcal{R}(\mathcal{V}_L) \rightarrow \mathcal{R}(\mathcal{V}_H)$, an *intervention distribution* $\Pr_{\mathcal{I}}$ is a distribution on $\mathcal{I}_L \times \mathcal{I}_H$ such that $\Pr_{\mathcal{I}}(\vec{X} \leftarrow \vec{x} \mid \vec{Y} \leftarrow \vec{y}) > 0$ iff $\omega_\tau(\vec{X} \leftarrow \vec{x}) = \vec{Y} \leftarrow \vec{y}$. ∎

We think of $\Pr_{\mathcal{I}}(\vec{X} \leftarrow \vec{x} \mid \vec{Y} \leftarrow \vec{y})$ as telling us how likely the high-level intervention $\vec{Y} \leftarrow \vec{y}$ is to have been implemented by the low-level intervention $\vec{X} \leftarrow \vec{x}$.

Definition 4.8: Given a surjective map $\tau : \mathcal{R}(\mathcal{V}_L) \rightarrow$

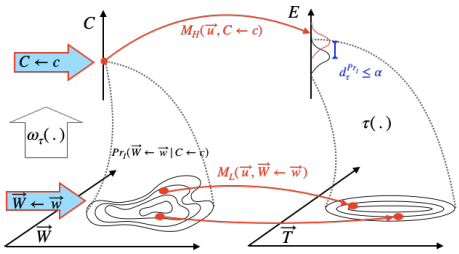

Figure 4: Probabilistic approximate abstraction for the climate example: An intervention on the high-level wind variable C results, given the distribution over contexts $\vec{u}$ (omitted for clarity), in the red distribution over the high-level temperature variable E. The intervention $C \leftarrow c$ can be instantiated in the low-level wind map $\vec{W}$ in many ways, according to the intervention distribution $Pr_\mathcal{I}$. Mapping the (manipulated) distribution over $\vec{W}$ using M_L results in a distribution over the low-level temperature map $\vec{T}$ that combines uncertainty from $Pr_\mathcal{I}$ and $Pr_\mathcal{U}$. Abstracting this distribution using τ to E results in the black distribution over E. M_H is a probabilistic τ-α-approximate abstraction of M_L if the distance between the expectations of these distributions is less than α.

$\mathcal{R}(\mathcal{V}_H)$ and an intervention-distribution $\Pr_\mathcal{I}$, the distance function $d_\tau^{\Pr_\mathcal{I}}$ on pairs of τ-consistent probabilistic causal models is defined by taking

$$d_\tau^{\Pr_\mathcal{I}}(M_L, M_H) = \min_{\{\tau_\mathcal{U}:\, \tau_\mathcal{U}(\Pr_L)=\Pr_H\}} \max_{\vec{Y}\leftarrow\vec{y}\in\mathcal{I}_H} \Big(\sum_{\vec{u}_L\in\mathcal{R}_L(\mathcal{U}_L)} \Pr_L(\vec{u}_L) \sum_{\vec{X}\leftarrow\vec{x}\in\mathcal{I}_L} \Pr_\mathcal{I}(\vec{X}\leftarrow\vec{x} \mid \vec{Y}\leftarrow\vec{y}) \, d_{\mathcal{V}_H}(\tau(M_L(\vec{u}_L, \vec{X}\leftarrow\vec{x})), M_H(\tau_\mathcal{U}(\vec{u}_L), \omega_\tau(\vec{X}\leftarrow\vec{x})))\Big).$$

∎

Intuitively, for each high-level intervention $\vec{Y} \leftarrow \vec{y}$, we take the expected distance between the two ways of lifting a low-level intervention to $\vec{Y} \leftarrow \vec{y}$. In computing the expectation, there are two sources of uncertainty: the likelihood of a given context (this is determined by $\Pr_L$) and the likelihood on each low-level intervention that maps to $\vec{Y} \leftarrow \vec{y}$ (this is given by $\Pr_\mathcal{I}(\cdot \mid \vec{Y} \leftarrow \vec{y})$). Fig. 4 illustrates the point for the climate example.

We can also define d_β^τ and $d_\beta^{\tau,\Pr_\mathcal{I}}$ in a manner completely analogous to Definition 4.2. We omit these definitions for reasons of space, but the details should be clear.

It is worth comparing our approach to other approaches to determining the distance between causal models. The more standard way to compare two causal models is to compare their causal graphs. The causal graph is a directed acyclic graph (dag) that has nodes labeled by variables; (the node labeled) X is an ancestor of (the node labeled) Y iff $X \prec Y$. Dags have been compared using what is called the *structural Hamming distance* (SHD) (Acid and de Campos 2003), where the SHD between G and H (which are assumed to have an identical set of nodes) is the number of pairs of nodes (i, j) on which G and H differ regarding the edge between i and j (either because one of one of them has an edge and the other does not, or the edges are oriented in different directions). As Peters and Bühlmann (2015) observe, the SHD misses out on some important information in causal networks. In particular, it does not really compare the effect of interventions. They want a notion of distance that takes this into account, as do we. However, they take into account the effect of interventions in a way much closer in spirit to SHD. Roughly speaking, in our language, given two similar causal models M_1 and M_2, they count the number of pairs (X, Y) of endogenous variables such that intervening on X leads to a different distribution over Y in M_1 and M_2. More formally, let $\Pr_Y^{M,X\leftarrow x}$ denote the marginal of $\Pr^{M,X\leftarrow x}$ on the variable Y in model M. (Since we want to compare probability distributions in two different models, we add the model to the superscript.) The SID between similar models M_1 and M_2 is the number of pairs (X, Y) such that there exists an $x \in \mathcal{R}(X)$ such that $\Pr_Y^{M_1,X\leftarrow x} \neq \Pr_Y^{M_2,X\leftarrow x}$.

Although SID does compare the predictions that two models make, it differs from our distance functions in several important respects. First, it compares just the effect of interventions on single variables, whereas we allow arbitrary interventions. We believe that it is important to consider arbitrary interventions, since sometimes variables act together, and it takes intervening on more than one variable to distinguish two models. Second, we are interested in how far apart two distributions are, not just the fact that they are different. Finally, we want a definition that applies to models that are not similar, since this is what we need for approximate abstraction.

5 COMPOSING ABSTRACTION AND APPROXIMATION

It can be useful to understand an approximate abstraction as the result of composing an approximation and an abstraction, in some order. For example, we can explain the ideal gas law in terms of thinking of frictionless elastic collisions between particles (this is an approximation to the truth) and then abstracting by replacing the kinetic energy of the particles by their temperature (a measure of average kinetic energy). Here we examine the extent to which an approximate abstraction can be viewed this way. We start with two easy results showing that if we compose an approximation and an abstraction in some order, then we do get an approximate abstraction.

Proposition 5.1 : *If $(M_H, \mathcal{I}_H)$ is a τ-abstraction of $(M_L, \mathcal{I}_L)$ and $(M'_H, \mathcal{I}_H)$ is a d_{max}-α-approximation of $(M_H, \mathcal{I}_H)$ then $(M'_H, \mathcal{I}_H)$ is a τ-α approximate abstraction of $(M_L, \mathcal{I}_L)$.* [4]

In Proposition 5.1 we considered an abstraction followed by an approximation. Things change if we do things in the opposite order; that is, if we consider an approximation followed by an abstraction. For suppose that M'_L is a d_{max}-α approximation of M_L and M_H is a τ-abstraction of M_L. Now it is not in general the case that M_H is a d_τ-α-approximate abstraction of M'_L. The problem is that when assessing how good an abstraction M_H is of M'_L, we are comparing two high-level states (using $d_{\mathcal{V}_H}$). But when comparing M_L to M'_L, we use $d_{\mathcal{V}_L}$. In general, $d_{\mathcal{V}_L}(\vec{v}_L, \vec{v}'_L)$ and $d_{\mathcal{V}_H}(\tau(\vec{v}_L), \tau(\vec{v}'_L))$ may be unrelated.

Proposition 5.2 : *If $(M_H, \mathcal{I}_H)$ is a τ-abstraction of $(M_L, \mathcal{I}_L)$ and $(M_L, \mathcal{I}_L)$ is a d_{max}-α approximation of $(M'_L, \mathcal{I}_L)$, then $(M_H, \mathcal{I}_H)$ is a τ-$k\alpha$ approximate abstraction of $(M'_L, \mathcal{I}'_L)$, where k is*

$$\max_{\vec{X} \leftarrow \vec{x} \in \mathcal{I}_L, \vec{u}_L \in \mathcal{R}_L(\mathcal{U}_L)} \frac{d_{\mathcal{V}_H}(\tau(M'_L(\vec{u}_L, \vec{X} \leftarrow \vec{x})), \tau(M_L(\vec{u}_L, \vec{X} \leftarrow \vec{x})))}{d_{\mathcal{V}_L}(M'_L(\vec{u}_L, \vec{X} \leftarrow \vec{x}), M_L(\vec{u}_L, \vec{X} \leftarrow \vec{x}))}.$$

While composing an approximation and an abstraction gives us an approximate abstraction, the full paper (2019) contains two examples that show that we cannot in general decompose an approximate abstraction into an abstraction composed with an approximation or an approximation composed with an abstraction.

As the following result shows, we can test whether an approximate abstraction can be viewed as the result of composing an abstraction and an approximation.

Theorem 5.3: *If $(M_H, \mathcal{I}_H)$ is a τ-α-approximate abstraction of $(M_L, \mathcal{I}_L)$ then the problem of determining whether there exists a model $(M'_H, \mathcal{I}_H)$ (resp., $(M'_L, \mathcal{I}_L)$) that is similar to $(M_H, \mathcal{I}_H)$ and is an abstraction of $(M_L, \mathcal{I}_L)$ (resp., $(M_L, \mathcal{I}_L)$) is in nondeterministic time polynomial in $|\mathcal{I}_L| \times |\mathcal{R}_L(\mathcal{U}_L)|$.*

There is one important special case where we are guaranteed to be able to find an appropriate intermediate low-level model, and to do so in polynomial time: if the mapping τ is constructive.

Theorem 5.4: *If $\tau : \mathcal{R}_L(\mathcal{V}_L) \to \mathcal{R}_H(\mathcal{V}_H)$ is constructive, the causal models $(M_L, \mathcal{I}_L)$ and $(M_H, \mathcal{I}_H)$ are τ-consistent, and $(M_H, \mathcal{I}_H)$ is a τ-α-approximate abstraction of $(M_L, \mathcal{I}_L)$, then we can find a model $(M'_L, \mathcal{I}_L)$ that is similar to $(M_L, \mathcal{I}_L)$, and such that $(M_H, \mathcal{I}_H)$ is a τ-abstraction of $(M'_L, \mathcal{I}_L)$ in time polynomial in $|\mathcal{I}_L| \times |\mathcal{R}'_L(\mathcal{U}_L)|$.*

The full paper contains an example showing that τ being constructive does not similarly guarantee the existence of M'_H in the first half of Theorem 5.3.

[4] Proofs of all technical results can be found in the full paper (Beckers, Eberhardt, and Halpern 2019).

6 DISCUSSION AND CONCLUSIONS

By defining notions of abstraction, approximation, and approximate abstraction, we have presented a framework that relates causal models that describe the same system at (possibly) different levels of granularity. While coarser models offer a degree of simplification by omitting details, they also in general entail a loss in accuracy with respect to the fundamental description. Our framework shows how to quantify this loss in accuracy by defining a distance metric that captures the degree to which a more abstract causal model approximates a more detailed causal model. High- and low-level causal models of the same system can vary on almost any dimension. They need not share the same equations, the same variables, or the same interventions. They may involve entirely distinct state spaces. Inevitably, then, there is some degree of choice as to what one deems relevant to the approximation.

Starting with deterministic causal models, we provided a general method for quantifying the "goodness" of an approximate abstraction. As an interesting special case, our approach allows for the comparison of causal models that operate at the same level of detail. We then extended to probabilistic causal models, and considered several different choices for quantifying the distance between models. Finally, we considered the extent to which we could decompose an approximate abstraction into an abstraction and an approximation.

Given the ubiquitous use of causal models in the social and natural sciences that are known not to capture all the causally relevant details, the framework we presented offers a principled way to assess the trade-off between abstraction and accuracy.

Acknowledgements

Beckers was supported by the grant ERC-2013- CoG project REINS 616512. Eberhardt was supported in part by NSF grants 1564330 and BCS-1845958, and HRGC grant 13601017. Halpern was supported in part by NSF grants IIS-1703846 and IIS-1718108, ARO grant W911NF-17-1-0592, and a grant from the Open Philanthropy project. We thank the UAI reviewers for many useful comments.

References

Acid, S. and L. M. de Campos (2003). Searching for Bayesian network structures in the space of restricted acyclic partially directed graphs. *Journal of A.I. Research 18*, 445–490.

Beckers, S., F. Eberhardt, and J. Y. Halpern (2019). Approximate causal models. Available at arxiv.org/abs/1906.11583.

Beckers, S. and J. Y. Halpern (2019). Abstracting causal models. In *Proc. Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*. The full version is available at arxiv.org/abs/1812.03789.

Boutilier, C., N. Friedman, M. Goldszmidt, and D. Koller (1996). Context-specific independence in Bayesian networks. In *Proc. Twelfth Conference on Uncertainty in Artificial Intelligence (UAI '96)*, pp. 115–123.

Chalupka, K., T. Bischoff, P. Perona, and F. Eberhardt (2016). Unsupervised discovery of El Niño using causal feature learning on microlevel climate data. In *Proc. 32nd Conference on Uncertainty in Artificial Intelligence (UAI 2016)*, pp. 72–81.

Halpern, J. Y. (2016). *Actual Causality*. Cambridge, MA: MIT Press.

Halpern, J. Y. and J. Pearl (2005). Causes and explanations: a structural-model approach. Part I: Causes. *British Journal for Philosophy of Science 56*(4), 843–887.

Peters, J. and P. Bühlmann (2015). Structural intervention distance (SID) for evbaluating causal graphs. *Neural Computation 27*(3), 725–747.

Rubenstein, P. K., S. Weichwald, S. Bongers, J. M. Mooij, D. Janzing, M. Grosse-Wentrup, and B. Schölkopf (2017). Causal consistency of structural equation models. In *Proc. 33rd Conference on Uncertainty in Artificial Intelligence (UAI 2017)*.

Spirtes, P. and R. Scheines (2004). Causal inference of ambiguous manipulations. *Philosophy of Science 71*, 833–845.