Combining the Causal Judgments of Experts with Possibly Different Focus Areas

Meir Friedenberg
Department of Computer Science
Cornell University
meir@cs.cornell.edu

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

Abstract

In many real-world settings, a decision-maker must combine information provided by different experts in order to decide on an effective policy. Alrajeh, Chockler, and Halpern (2018) showed how to combine causal models that are compatible in the sense that, for variables that appear in both models, the experts agree on the causal structure. In this work we show how causal models can be combined in cases where the experts might disagree on the causal structure for variables that appear in both models due to having different focus areas. We provide a new formal definition of compatibility of models in this setting and show how compatible models can be combined. We also consider the complexity of determining whether models are compatible. We believe that the notions defined in this work are of direct relevance to many practical decision making scenarios that come up in natural, social, and medical science settings.

1 Introduction

In many real-world settings, a decision-maker must combine information provided by different experts in order to decide on an effective policy. For example, when deciding policing and criminal justice policy, it may be necessary to consult different experts specializing in areas such as criminology, psychology, sociology, and economics. Intelligently combining the information provided by the various experts is necessary if the decision-maker hopes to select the best course of action.

Much work has been done on combining simple probabilistic judgments of different experts. However, we are interested in settings where a decision-maker wants to choose an action in order to induce a particular outcome, so we are interested in the setting where experts provide models of the *causal* relationships between different factors. Despite the clear importance of combining causal models in real-world situations, there has been very little work on how to combine models with this extra structure.

Much work has been done on the related problem of learning causal models: given data and possibly some prior structured knowledge, extract the causal model that best fits the given information (see, e.g., (Claassen and Heskes 2010; 2012; Hyttinen, Eberhardt, and Jarvisalo 2014; Tillman and Spirtes 2011; Triantafillou and Tsamardinos 2015); Triantafillou and Tsamardinos (2015) provide a good overview of work in the area). In a certain sense, if all of the experts we consult have learned their models in this manner and can provide us with all of their data, then the best thing to do is simply to learn a new causal model from the union of all the data. However, in many real world settings, this is completely impractical. Experts develop intuitions based on years worth of data, training, and discussions; providing all of this background information to the decision maker may be infeasible.

On the topic of combining causal models without data, Bradley, Dietrich, and List (2014) proved an impossibility result. Given a set of desiderata for combining causal models, they show that no algorithm satisfies them all. They further examine ways of circumventing the impossibility result by weakening some of those conditions.

It is perhaps not too surprising in retrospect that it will sometimes be impossible to combine causal models, as two models can explicitly disagree on every causal relationship. In the work most related to this, Alrajeh, Chockler, and Halpern (2018) (ACH from now on) provide conditions for the *compatibility* of models and show how to combine models that meet their compatibility conditions. They define a dominance relation according to which a model M_1 dominates model M_2 with respect to a variable C if the two models agree on the causal dependence of C on the other variables shared by the two models, but model M_1 has perhaps a more detailed picture of the exact way that the effects are mediated. Two models are compatible if, for every variable, one of the models dominates the other; the combined model takes the causal information from the dominant model for each variable C. ACH also provide a way of assigning probabilities to causal models in settings where not all models under consideration are compatible.

The present work can be seen as providing an approach complementary to that of ACH. Philosophically, the approach presented by ACH is intended to allow for combination of models where the modelers fundamentally agree on the causal relationship between the variables they both discuss, but go into different levels of detail as to how some of those relationships are mediated. But consider, for instance, the following scenario: a medical scientist is interested in the conditions under which a particular reaction occurs, and

consults with two experts. The first specializes in the exact mechanism by which this reaction occurs; the second specializes in how one of the reactants gets produced. Because of their different focus areas, they in fact do not agree on everything; each of them is more aware of the details of the reaction that she studies, and thus has more understanding of what factors can cause that reaction to occur differently. Our intuition tells us that there should be a way of combining these models to get the true expertise of both modelers, but with the ACH approach, these models would in fact have to be deemed incompatible.

In this work, we allow for combining models where the modelers disagree due to their different focus areas. Intuitively, if the first modeler considered more possibilities than the second, and her conclusion can *explain* the observations of the second, then we accept the conclusion of the first modeler. To this end, we define a new notion of a *"can explain"* relation and provide new formalizations of compatibility and model combination relative to this notion.

The rest of this paper is organized as follows. In Section 2, we review the basic framework of causal models, and extend them so as to accommodate focus areas. In Section 3, we define our approach to combining these models. Section 4 contains an approach to weighting models in settings where the models under consideration are not all compatible. We characterize the computational complexity of the can-explain relation that we define in Section 5. Section 6 concludes.

2 Causal Models with Focus

In this section, we review the framework of causal models. We largely follow Halpern and Pearl (**?**), but extend their basic framework so as to allow the models to express focus areas.

We assume that a situation is characterized by the values of a number of variables. There are *structural equations* describing the effect that the variables have on each other. Among the variables, we distinguish between *exogenous variables* (whose values are determined by factors outside of the model) and *endogenous variables* (whose values are determined by other variables in the model).

A *causal model with focus* is a tuple $M = (\mathcal{S}, \mathcal{F}, \mathcal{G})$, where $\mathcal{S}$ is a *signature*, $\mathcal{F}$ is a set of *structural equations*, and $\mathcal{G}$ is a *focus function*. The signature $\mathcal{S}$ is itself a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{R})$. Here $\mathcal{U}$ is a (finite but non-empty) set of exogenous variables and $\mathcal{V}$ is a (finite but non-empty) set of endogenous variables. $\mathcal{R}$ is a *range* function mapping elements of $\mathcal{U} \cup \mathcal{V}$ to the (finite but non-empty) set of values they can take on. We assume without loss of generality that $|\mathcal{R}(C)| > 1$ for all variables C. (If a variable can take on only one value, then it can neither be a cause nor have its value be caused by another variable, so we can remove it and get a semantically equivalent model.) $\mathcal{F}$ associates with each endogenous variable $X \in \mathcal{V}$ a function denoted F_X such that $F_X : (\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V} - \{X\}} \mathcal{R}(Y)) \rightarrow \mathcal{R}(X)$; that is, F_X determines the value of X, given the values of all the other variables in $\mathcal{U} \cup \mathcal{V}$. For example, we might have $F_X(u, y, z) = u+y$, which is usually written as $X = U+Y$. Thus, if $Y = 3$ and $U = 2$, then $X = 5$, regardless of how Z is set.

Up to now, we have essentially described the Halpern-Pearl (**?**) model. In our setting, though, we add an additional *focus function* that intuitively tells us what variables the modeler considered when trying to determine the structural equation for each variable. In practice, we might extract such information from the modeler herself or from the published experiments of the modeler. Formally, we let $\mathcal{G} : (\mathcal{U} \cup \mathcal{V}) \rightarrow 2^{(\mathcal{U} \cup \mathcal{V})}$ be a function that, given a variable C, gives us the set of variables that the modeler considered as possibly having an effect on C. We require for all C that $C \notin \mathcal{G}(C)$, as a variable cannot have a causal effect on itself. It may seem surprising at first that we define this function even for exogenous variables, which are not affected by other variables in the model. We think that it is more natural to do so, as it is possible that the modeler considered the possibility of the variable being endogenous before deciding that it wasn't affected by any other variables in the model.

We define B to be a *parent* of C if there exists some setting of all the variables in $(\mathcal{V} \cup \mathcal{U}) - \{B, C\}$ such that C takes on some value c_1 for a value b_1 of B and takes on a different value c_2 for some other values b_2 of B. Let $Par(C)$ be the set of parents of C. Thus, the parents of C are exactly those variables that might have a direct effect on C. We require that $Par(C) \subseteq \mathcal{G}(C)$ for every variable C. A modeler cannot have an equation for C showing that B has a direct influence on C unless the modeler considered B as a possibly having an effect on C.

A causal model with focus with exogenous variables $\mathcal{U}$ and endogenous variables $\mathcal{V}$ can be represented by a pair of graphs on $\mathcal{U} \cup \mathcal{V}$. In the first graph, called the *parent graph*, the edge set E consists of edges from the vertices in $Par(C)$ to C, for each endogenous variable C. In the second graph, called the *focus graph*, the edge set E' consists of edges to each vertex C from the members of $\mathcal{G}(C)$. Pictorially, we can depict this representation with directed edges for the elements of E and crossed-out directed edges for the elements of $E' - E$. We call a model *recursive* or *acyclic* if the parent graph does not contain any cycles. In cases where the model is acyclic, given a context $\vec{u}$ (i.e., a setting of the exogenous variables), the values of all the endogenous variables are uniquely determined by the structural equations. As is standard in the literature, we restrict our discussion to acyclic models in this work.

Given a model M, an endogenous variable $X \in \mathcal{V}$, and a value $x \in \mathcal{R}(X)$, we define $M_{X \leftarrow x}$ to be the model that is the same as M except that the equation for X is replaced by $X = x$. We can think of the model $M_{X \leftarrow x}$ as describing the result of intervening to set X to x in model M.

Take a *causal formula* to be one of the form $[Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi$, where $Y_j \in \mathcal{U} \cup \mathcal{V}$ and φ is a Boolean combination of *primitive formulas* of the form $X = x$, where X is an endogenous variable and $x \in \mathcal{R}(X)$.[1] In the special case where $k = 0$, we identify $[]\varphi$ with the formula φ.

We now define what it means for a causal formula φ to be true in a *causal setting* $(M, \vec{u})$ consisting of a causal

[1] In previous work, each Y_i is taken to be an endogenous variable. For our purposes, it is useful to also allow Y to be exogenous.

model M and a context $\vec{u}$, written $(M, \vec{u}) \models \varphi$, by induction on the structure of φ. For a primitive event $X = x$, $(M, \vec{u}) \models X = x$ if $X = x$ in the unique solution to the equations in M given context $\vec{u}$ (the solution is unique since we are dealing with acyclic models, so the setting of the exogenous variables determines all other variables). The truth of a Boolean combination of primitive events is defined in the obvious way. If $k \geq 1$ and Y_k is an endogenous variable, then

$$(M, \vec{u}) \models [Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi \text{ iff}$$
$$(M_{Y_k \leftarrow y_k}, \vec{u}) \models [Y_1 \leftarrow y_1, \ldots, Y_{k-1} \leftarrow y_{k-1}]\varphi.$$

If Y_k is an exogenous variable, then

$$(M, \vec{u}) \models [Y_1 \leftarrow y_1, \ldots, Y_k \leftarrow y_k]\varphi \text{ iff}$$
$$(M, \vec{u}[Y_k/y_k]) \models [Y_1 \leftarrow y_1, \ldots, Y_{k-1} \leftarrow y_{k-1}]\varphi,$$

where $\vec{u}[Y_k/y_k]$ is the result of replacing the value of Y_k in $\vec{u}$ by y_k.

We now show how to model the example from the introduction in this framework.

Example 2.1. Recall the basic scenario: a medical scientist is trying to understand under what conditions a particular reaction occurs, and consults with two experts. The first specializes in the exact mechanism by which this reaction occurs; the second in how one of the reactants gets produced. The scientist then wants to combine the information provided by the two experts. The models provided by the experts are depicted in Figure 1, where expert i provides model M_i. The main difference between these models is that expert 1 takes into account the effect that temperature T can have on reaction C, while modeler 2, who does not, takes into account the effect temperature can have on the production of reactant B. Formally, the parameters of these two models are defined as follows: for the ranges, we have $\mathcal{R}_1(T) = \mathcal{R}_2(T) = \{\mathsf{Freezing}, \mathsf{Cool}, \mathsf{Hot}\}$, $\mathcal{R}_1(A') = \mathcal{R}_1(B') = \mathcal{R}_2(A') = \mathcal{R}_2(B') = \{1, 2, 3\}$, $\mathcal{R}_1(A) = \mathcal{R}_1(B) = \mathcal{R}_2(A) = \mathcal{R}_2(B) = \{1, \ldots, 5\}$, and $\mathcal{R}_1(C) = \mathcal{R}_2(C) = \{\mathsf{true}, \mathsf{false}\}$. We have $\mathcal{G}_1(A) = \{A'\}$, $\mathcal{G}_1(B) = \{B'\}$, and $\mathcal{G}_1(C) = \{A, B, T\}$, while $\mathcal{G}_1(A) = \{A'\}$, $\mathcal{G}_1(B) = \{B', T\}$, and $\mathcal{G}_1(C) = \{A, B\}$. This is how we model the fact that expert 1 does not take into account the effect that temperature (T) can have on B, while expert 2 does not take into account the effect that temperature can have on C. The structural equations in M_1 are defined by taking $A = A'$, $B = B'$, and $C = ((T = \mathsf{Freezing}) \wedge (A + B \geq 9)) \vee ((T = \mathsf{Cool}) \wedge (A + B \geq 5)) \vee ((T = \mathsf{Hot}) \wedge (A + B \geq 4))$. In M_2, the structural equations are $A = A'$; $B = B' + 2$ if $T = \mathsf{Freezing}$ and $B = B'$ otherwise; and $C = \mathsf{true}$ if $A + B \geq 5$ and $C = \mathsf{false}$ otherwise. □

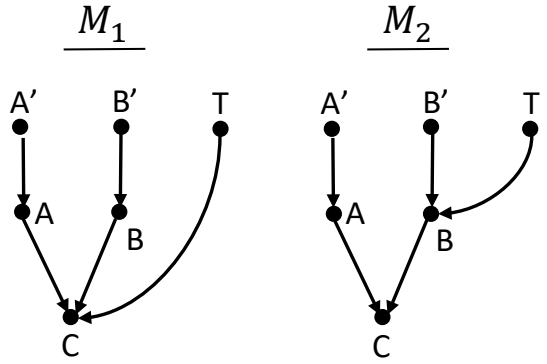

Figure 1: M_1 and M_2 are the models of the two scientists trying to understand reaction C.

3 Combining Causal Models with Focus

In this section we turn to the question of combining causal models. We define a new relation and show how it can be used to define compatibility and combination.

3.1 The "can-explain" relation

We want to capture the intuition that if modeler i considered the causes of some variable C more carefully than modeler j, then i's analysis is preferable. Roughly speaking, we prefer modeler i's structural equation for C over j's if i's model can explain (in some appropriate sense) j's observations.

Before going on, we introduce some notation conventions that will simplify the exposition. When we write M_i, we assume that the model M_i has components $((\mathcal{U}_i, \mathcal{V}_i, \mathcal{R}_i), \mathcal{F}_i, \mathcal{G}_i)$, and $Par_i(C)$ refers to the parents of variable C in model M_i. Also, for a model M, we write $\mathcal{F}^M$ to denote the $\mathcal{F}$ function in model M, and similarly for the other components of the model.

Definition 3.1. M_1 *can explain* M_2 with respect to C, written $M_1 \succeq_C M_2$, if

(a) $\mathcal{R}_1(C) = \mathcal{R}_2(C)$,

(b) $\mathcal{G}_2(C) \subseteq \mathcal{G}_1(C)$, and

(c) for all exogenous settings $\vec{u}_2$ for M_2 and all interventions $\mathcal{G}_2(C) = \vec{x}$ there is a context $\vec{u}_1$ in M_1 such that if $(M_2, \vec{u}_2) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$ then $(M_1, \vec{u}_1) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$.

This relation ought really be called *can explain and has considered everything considered by*, but for the sake of brevity we use simply *can explain*.

The intuition here is that expert 2 (whose knowledge is characterized by M_2) has considered carefully the effect on C of all the variables in $\mathcal{G}_2(C)$ and has observed that those in $Par_2(C)$ have an effect on C, while those in $\mathcal{G}_2(C) - Par_2(C)$ do not. She has not bothered considering the effect of the variables not in $\mathcal{G}_2(C)$ on C, because she is reasonably sure that they have no effect (but could turn out to be wrong about this). Expert 1 (whose knowledge is characterized by M_1) can explain expert 2's observations (at least, with regard to C) if she has also considered at least all of the interventions that expert 2 has considered, and can explain all of expert 2's observations in the sense of condition (c) of Definition 3.1.

We conclude this subsection with two technical results that highlight useful properties of the $\succeq_C$ relation. Say that M_1 and M_2 are *C-compatible* if either $M_1 \succeq_C M_2$ or $M_2 \succeq M_1$. We now show that $\succeq_C$ is transitive when restricted to C-compatible models.

Proposition 3.2. *If $M_1 \succeq_C M_2$, $M_2 \succeq_C M_3$, and M_1 and M_3 are C-compatible, then $M_1 \succeq_C M_3$.*

Proof. Assume that $M_1 \succeq_C M_2$, $M_2 \succeq_C M_3$, and by way of contradiction, that $M_1 \not\succeq_C M_3$. Because M_1 and M_3 are C-compatible, it must then be the case that $M_3 \succeq_C M_1$. Since $\mathcal{R}_1(C) = \mathcal{R}_2(C)$ and $\mathcal{R}_2(C) = \mathcal{R}_3(C)$, we have that $\mathcal{R}_1(C) = \mathcal{R}_3(C)$. And since $\mathcal{G}_2(C) \subseteq \mathcal{G}_1(C)$, $\mathcal{G}_3(C) \subseteq \mathcal{G}_2(C)$, and $\mathcal{G}_1(C) \subseteq \mathcal{G}_3(C)$, we get that $\mathcal{G}_1(C) = \mathcal{G}_2(C) = \mathcal{G}_3(C)$. Now consider any intervention $\mathcal{G}_3(C) = \vec{x}$. Because $M_2 \succeq_C M_3$, we know that any value of C that can be achieved in M_3 under intervention $\vec{X} = \vec{x}$ can also be achieved in M_2 under the same intervention; that is, for all contexts $\vec{u}$ and values $c \in \mathcal{R}(C)$, if $(M_3, \vec{u}) \models [\mathcal{G}_3(C) \leftarrow \vec{x}](C = c)$, then there exists a context $\vec{u}'$ such that $(M_2, \vec{u}') \models [\mathcal{G}_3(C) \leftarrow \vec{x}](C = c)$. But because $\mathcal{G}_2(C) = \mathcal{G}_3(C)$ and $M_1 \succeq_C M_2$, it follows that there exists a context $\vec{u}''$ such that $(M_1, \vec{u}'') \models [\mathcal{G}_3(C) \leftarrow \vec{x}](C = c)$. Thus, condition (c) of Definition 3.1 holds, so $M_1 \succeq_C M_3$. □

The requirement in Proposition 3.2 that M_1 and M_3 are C-compatible is necessary, as we show below (see Example 3.8).

Definition 3.3. $M_1 \equiv_C M_2$ iff either (a) $C \in U_1 \cap U_2$, $\mathcal{R}_1(C) = \mathcal{R}_2(C)$, and $\mathcal{G}_1(C) = \mathcal{G}_2(C)$ or (b) $C \in V_1 \cap V_2$, $\mathcal{R}_1(C) = \mathcal{R}_2(C)$, $\mathcal{G}_1(C) = \mathcal{G}_2(C)$, and $\mathcal{F}_1(C) = \mathcal{F}_2(C)$.[2]

The next result shows that, in a sense, $\succeq_C$ is antisymmetric.

Proposition 3.4. *$M_1 \succeq_C M_2$ and $M_2 \succeq_C M_1$ iff $M_1 \equiv_C M_2$.*

Proof. The fact that $M_1 \equiv_C M_2$ implies $M_1 \succeq_C M_2$ and $M_2 \succeq_C M_1$ follows easily from the definitions, using the fact that $Par_i(C) \subseteq \mathcal{G}_i(C)$.

To prove the opposite implication, suppose that $M_1 \succeq_C M_2$ and $M_2 \succeq_C M_1$. We first show that C cannot be in either $\mathcal{U}_1 \cap \mathcal{V}_2$ or $\mathcal{U}_2 \cap \mathcal{V}_1$. Suppose, by way of contradiction, that $C \in \mathcal{U}_1 \cap \mathcal{V}_2$. Consider an intervention $\mathcal{G}_2(C) = \vec{x}$. Because C is exogenous in M_2, there must exist contexts $\vec{u}_2 \neq \vec{u'}_2$ such that $(M_2, \vec{u}_2) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c_2)$ and $(M_2, \vec{u}'_2) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c'_2)$ for some c_2 and c'_2 such that $c_2 \neq c'_2$. Now consider this same intervention in M_1. Since $M_1 \succeq_C M_2$ and $M_2 \succeq M_1$, we have that $\mathcal{G}_1(C) = \mathcal{G}_2(C)$. By definition, $Par_1(C) \subseteq \mathcal{G}_1(C)$. Thus there must exist a unique c_1 such that, for all exogenous settings $\vec{u}_1$ in M_1, $(M_1, \vec{u}_1) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c_1)$. But because $c_2 \neq c'_2$, there cannot be contexts $\vec{u}_1$ and $\vec{u}'_1$ such that $(M_1, \vec{u}_1) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c_2)$ and $(M_1, \vec{u}'_1) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c'_2)$. This contradicts the assumption that $M_1 \succeq_C M_2$. A similar argument shows that C cannot be in $\mathcal{U}_2 \cap \mathcal{V}_1$.

It is almost immediate from the definition of $\succeq_C$ that if $M_1 \succeq_C M_2$, $M_2 \succeq_C M_1$, and $C \in (\mathcal{U}_1 \cap \mathcal{U}_2) \cup (\mathcal{V}_1 \cap \mathcal{V}_2)$, then $\mathcal{R}_1(C) = \mathcal{R}_2(C)$ and $\mathcal{G}_1(C) = \mathcal{G}_2(C)$. It follows that if $C \in \mathcal{U}_1 \cap \mathcal{U}_2$, then $M_1 \equiv_C M_2$. It remains to show that if $C \in \mathcal{V}_1 \cap \mathcal{V}_2$, then $\mathcal{F}_1(C) = \mathcal{F}_2(C)$.

So suppose that $C \in V_1 \cap V_2$. If $Par_1(C) = Par_2(C)$, then since $\mathcal{G}_1(C) = \mathcal{G}_2(C)$ and $Par_i(C) \subseteq \mathcal{G}_i(C)$ for $i = 1, 2$, it follows that $(M_1, \vec{u}_1) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$ iff $(M_2, \vec{u}_2) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$ for all contexts $\vec{u}_1$ and $\vec{u}_2$, so $\mathcal{F}_1(C) = \mathcal{F}_2(C)$. On the other hand, if $Par_1(C) \neq Par_2(C)$, then without loss of generality there is some variable $D \in Par_1(C) - Par_2(C)$. There must thus exist two interventions $\mathcal{G}_2(C) = \vec{x}$ and $\mathcal{G}_2(C) = \vec{y}$ that differ only on the value of D such that for some $c \in \mathcal{R}(C)$, we have $(M_1, u'_1) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$ and $(M_1, u'_1) \models [\mathcal{G}_2(C) \leftarrow \vec{y}]\neg(C = c)$ for all exogenous settings u'_1 in M_1. Because $D \notin Par_2(C)$, we know that interventions $\mathcal{G}_2(C) = \vec{x}$ and $\mathcal{G}_2(C) = \vec{y}$ will give the same value of C in M_2 for all settings of exogenous variables u_2. Thus, it is not the case that M_1 can explain M_2 with respect to C, giving a contradiction.

So we have in all cases that $M_1 \equiv_C M_2$, as desired. □

3.2 Combining compatible models

We now turn to compatibility and combination of causal models. We start by defining a simplified notion of compatibility and an operator $\oplus'$ that gets us most of the way there. Unfortunately, as we show, $\oplus'$ has a small shortcoming, so we then modify it to get a more reasonable operator $\oplus$.

Definition 3.5. *M_1 and M_2 are* compatible *if, for all $C \in (\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2)$, either $M_1 \succeq_C M_2$ or $M_2 \succeq_C M_1$.*

If M_1 and M_2 are compatible then, for each variable C, we intuitively want the combined model to take all of the information for C from the model that best explains C. So if M_1 can explain M_2 with respect to C, then we want the combined model to use M_1's focus function and structural equation (if C is endogenous in M_1) for C. Formally, the combined model $M_1 \oplus' M_2 = ((\mathcal{U}, \mathcal{V}, \mathcal{R}), \mathcal{F}, \mathcal{G})$ is defined as follows:

- $\mathcal{U} \cup \mathcal{V} = (\mathcal{U}_1 \cup \mathcal{V}_1) \cup (\mathcal{U}_2 \cup \mathcal{V}_2)$ (so the exogenous and endogenous variables in the combined model comprise all the endogenous and exogenous variables in M_1 and M_2). A variable U is exogenous in $M_1 \oplus' M_2$ if it is exogenous in one of M_1 or M_2, say M_i, and either does not appear in M_{3-i} (i.e., the other model) or it appears in M_{3-i} but $M_i \succeq_U M_{3-i}$; the remaining variables are endogenous. Formally, $\mathcal{U} = (\mathcal{U}_1 - (\mathcal{U}_2 \cup \mathcal{V}_2)) \cup (\mathcal{U}_2 - (\mathcal{U}_1 \cup \mathcal{V}_1)) \cup \{C : \exists i \in \{1, 2\}(C \in U_i \text{ and } M_i \succeq_C M_{3-i}\}$ and $\mathcal{V} = (\mathcal{V}_1 - (\mathcal{U}_2 \cup \mathcal{V}_2))) \cup (\mathcal{V}_2 - (\mathcal{U}_1 \cup \mathcal{V}_1)) \cup \{C : \exists i \in \{1, 2\}(C \in V_i \text{ and } M_i \succeq_C M_{3-i})\}$.
- For $C \in (\mathcal{U}_1 \cup \mathcal{V}_1) - (\mathcal{U}_2 \cup \mathcal{V}_2)$, set $\mathcal{R}(C) = \mathcal{R}_1(C)$, $\mathcal{F}(C) = \mathcal{F}_1(C)$, and $\mathcal{G}(C) = \mathcal{G}_1(C)$.
- Similarly, for $C \in (\mathcal{U}_2 \cup \mathcal{V}_2) - (\mathcal{U}_1 \cup \mathcal{V}_1)$, set $\mathcal{R}(C) = \mathcal{R}_2(C)$, $\mathcal{F}_C = \mathcal{F}_2(C)$, and $\mathcal{G}(C) = \mathcal{G}_2(C)$.
- For $C \in (\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2)$, we must have either $M_1 \succeq_C M_2$ or $M_2 \succeq_C M_1$. If $M_i \succeq M_{3-i}$, then set $\mathcal{R}(C) = \mathcal{R}_i(C)$, $\mathcal{F}(C) = \mathcal{F}_i(C)$, and $\mathcal{G}(C) = \mathcal{G}_i(C)$. (By Proposition 3.4, this is well defined: if $M_1 \succeq_C M_2$ and $M_2 \succeq_C M_1$, then $\mathcal{R}_1(C) = \mathcal{R}_2(C)$, $\mathcal{F}_1(C) = \mathcal{F}_2(C)$, and $\mathcal{G}_1(C) = \mathcal{G}_2(C)$.)

[2] Technically $\mathcal{F}_1(C) = \mathcal{F}_2(C)$ is not defined if $\mathcal{U}_1 \cup \mathcal{V}_1 \neq \mathcal{U}_2 \cup \mathcal{V}_2$; all we mean is that $Par_1(C) = Par_2(C)$, $\mathcal{R}_1(D) = \mathcal{R}_2(D)$ for all $D \in Par_1(C)$, and $(M_1, \vec{u}_1) \models [Par_1(C) \leftarrow \vec{p}](C = c)$ iff $(M_2, \vec{u}_2) \models [Par_2(C) \leftarrow \vec{p}](C = c)$ for all $\vec{u}_1, \vec{u}_2,$ and $\vec{p}$.

Returning to Example 2.1, it is easy to check that the models M_1 and M_2 are compatible. Specfically, we have $M_1 \succeq_C M_2$ and $M_2 \succeq_B M_1$. (For all other variables D, we have $M_1 \succeq_D M_2$ and $M_2 \succeq_D M_1$.) Thus, we can combine M_1 and M_2 to get the model $M_1 \oplus' M_2$ depicted in Figure 2

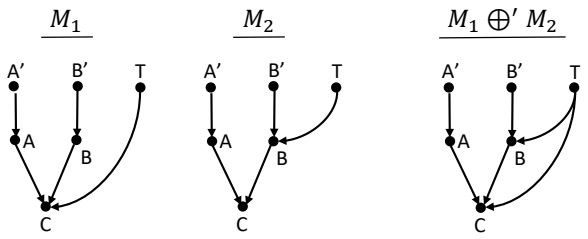

Figure 2: Taking into account what each scientist focused on allows us to combine M_1 and M_2 to get the model shown on the right.

In the ACH approach, M_1 and M_2 would be declared incompatible. Since no information about who considered what possibilities is available, expert 2 is assumed to have come to the conclusion that T does not directly affect C, and therefore be in fundamental disagreement with expert 1. In our setting, though, we can take advantage of the focus information to determine whether there is truly a fundamental disagreement. The disagreement may just be a result of the fact that one of the experts was not focusing on certain variables. In situations where this information is available, it can allow us to take more complete advantage of the different areas of expertise that different experts may have.

Slightly more generally, the ACH definition is designed to take into account situations where one expert's model is more detailed in terms of the topology of the causal graph, that is, where the causal relationship is considered to be mediated by variables the other modeler was simply not aware existed. In our setting there is more information available, allowing us to consider another sense in which one modeler's understanding might be locally more detailed than another's, namely, situations where one expert can explain the other's observed results by taking into account the fact that the other was not focusing on certain variables.

This notion of combination is commutative and, when defined, associative:

Proposition 3.6. *Given three pairwise compatible models M_1, M_2, and M_3,*

(a) $M_1 \oplus' M_2 = M_2 \oplus' M_1$;

(b) if $M_1 \succeq_C M_2$ *or* $C \in (\mathcal{U}_1 \cup \mathcal{V}_1) - (\mathcal{U}_2 \cup \mathcal{V}_2)$, *then* $M_1 \oplus' M_2 \equiv_C M_1$;

(c) if M_3 *is compatible with* $M_1 \oplus' M_2$ *and* M_1 *is compatible with* $M_2 \oplus' M_3$ *then* $M_1 \oplus' (M_2 \oplus' M_3) = (M_1 \oplus' M_2) \oplus' M_3$.

Proof. Commutativity is immediate from the definition of $\oplus'$.

For part (b), suppose that $M_1 \succeq_C M_2$. Then C is exogenous in $M_1 \oplus' M_2$ iff C is exogenous in M_1. Moreover, $\mathcal{R}^{M_1 \oplus' M_2}(C) = \mathcal{R}_1(C)$, $\mathcal{F}^{M_1 \oplus' M_2}(C) = \mathcal{F}_1(C)$ if $C \in \mathcal{V}_1$, and $\mathcal{G}^{M_1 \oplus' M_2}(C) = \mathcal{G}_1(C)$, so it immediately follows that $M_1 \oplus' M_2 \equiv_C M_1$. A similar argument applies if $C \in (\mathcal{U}_1 \cup \mathcal{V}_1) - (\mathcal{U}_2 \cup \mathcal{V}_2)$.

For part (c), observe that to show that $M_1 \oplus' (M_2 \oplus' M_3) = (M_1 \oplus' M_2) \oplus' M_3$, it suffices to show that $M_1 \oplus' (M_2 \oplus' M_3) \equiv_C (M_1 \oplus' M_2) \oplus' M_3$ for all $C \in (\mathcal{U}_1 \cup \mathcal{V}_1 \cup \mathcal{U}_2 \cup \mathcal{V}_2 \cup \mathcal{U}_3 \cup \mathcal{V}_3)$. We do this by considering, for each variable C, how many models it appears in.

First consider the case where C is in only one of the three models (i.e., $C \in U_i \cup \mathcal{V}_i$ for exactly one $i \in \{1, 2, 3\}$). Assume without loss of generality that C is in M_1. Then it follows almost immediately from our definitions that $M_1 \oplus' (M_2 \oplus' M_3) \equiv_C M_1 \equiv_C (M_1 \oplus' M_2) \oplus' M_3$, so $M_1 \oplus' (M_2 \oplus' M_3) \equiv_C (M_1 \oplus' M_2) \oplus' M_3$. Similarly, in the case where C is only in two models, assume without loss of generality that C is in M_1 and M_2. Then it follows immediately that $(M_1 \oplus' M_2) \oplus' M_3 \equiv_C M_1 \oplus' M_2$, so if $M_1 \succeq_C M_2$ then $(M_1 \oplus' M_2) \oplus' M_3 \equiv_C M_1$ and if $M_2 \succeq_C M_1$ then $(M_1 \oplus' M_2) \oplus' M_3 \equiv_C M_2$. It is also immediate that $M_2 \oplus' M_3 \equiv_C M_2$, so if $M_1 \succeq_C M_2$ then $M_1 \oplus' (M_2 \oplus' M_3) \equiv_C M_1$. Now consider the case where $M_2 \succeq_C M_1$. It must be the case that either $M_2 \oplus' M_3 \succeq_C M_1$ or $M_1 \succeq_C M_2 \oplus' M_3$ because they are compatible. If $M_2 \oplus' M_3 \succeq_C M_1$ then $M_1 \oplus' (M_2 \oplus' M_3) \equiv_C M_2$ and we are done. On the other hand, if $M_1 \succeq_C M_2 \oplus' M_3$ then, because $M_2 \oplus' M_3 \equiv_C M_2$, we know that $M_1 \succeq_C M_2$. But then because we assumed $M_2 \succeq_C M_1$ we get by Proposition 3.4 that $M_2 \equiv_C M_1$ and so $M_1 \oplus' (M_2 \oplus' M_3) \equiv_C M_1 \equiv_C M_2$. Thus, in al cases, we have that $(M_1 \oplus' M_2) \oplus' M_3 \equiv_C M_1 \oplus' (M_2 \oplus' M_3)$.

Finally, if C is in all three models, by Propositions 3.2 and 3.4, for some choice of i, j, k we have $M_i \succeq_C M_j \succeq M_k$. Suppose that $M_1 \succeq_C M_2 \succeq_C M_3$ (the argument is almost identical in all other cases). It follows from part (b) that $M_1 \oplus' M_2 \equiv_C M_1$. Because M_3 is compatible with $(M_1 \oplus' M_2)$, we know that either $(M_1 \oplus' M_2) \succeq_C M_3$ or $M_3 \succeq_C (M_1 \oplus' M_2)$. In the first case, it follows immediately from part (b) that $(M_1 \oplus' M_2) \oplus' M_3 \equiv_C M_1$. In the second case, since $(M_1 \oplus' M_2) \equiv_C M_1$ by part (b) and $M_3 \succeq_C (M_1 \oplus' M_2)$ by assumption, it follows that $M_3 \succeq_C M_1$. And since $M_1 \succeq_C M_2 \succeq_C M_3$, we have that $M_1 \succeq_C M_3$ by transitivity (Proposition 3.2), so it follows from Proposition 3.4 that $M_3 \equiv_C M_1$. But then from part (b) we have that $(M_1 \oplus' M_2) \oplus' M_3 \equiv_C M_3 \equiv_c M_1$. It is easy to show by similar reasoning that $M_1 \oplus' (M_2 \oplus' M_3) \equiv_C M_1$. So we get that $M_1 \oplus' (M_2 \oplus' M_3) \equiv_C (M_1 \oplus' M_2) \oplus' M_3$, completing the argument. □

One natural question to ask is whether this definition of combination is guaranteed to preserve acyclicity. Unfortunately, this is not the case, as the following example shows.

Example 3.7. Consider the models M_1 and M_2 in Figure 3, where

- $\mathcal{U}_1 = \{A, C\}$, $\mathcal{V}_1 = \{B, D\}$, $\mathcal{U}_2 = \{B, D\}$, and $\mathcal{V}_2 = \{A, C\}$;

- all variables are binary (i.e. have range $\{0, 1\}$);
- $\mathcal{G}_1(C) = \mathcal{G}_1(A) = \mathcal{G}_2(B) = \mathcal{G}_2(D) = \emptyset$, $\mathcal{G}_1(B) = \{A\}$, $\mathcal{G}_1(D) = \{C\}$, $\mathcal{G}_2(A) = \{D\}$, and $\mathcal{G}_2(C) = \{B\}$;
- in M_1, A is the parent of B and C is the parent of D, and in M_2, B is the parent of C and D is the parent of A. The details of the equations do not matter; for simplicity, suppose that in M_1 we have $B = A$ and $D = C$, while in M_2 we have $A = D$ and $C = B$.

Thus, A and C are exogenous in M_1, while B and D are exogenous in M_2. It is easy to see that, despite the fact that both models are acyclic, when we combine them we get a cyclic model. □

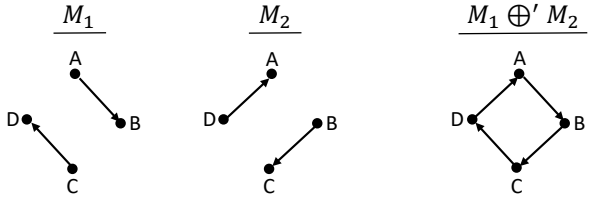

Figure 3: Although the models M_1 and M_2 are acyclic, the combined model $M_1 \oplus M_2$ contains a cycle.

We can, however, provide a simple and efficient test to guarantee that the combined model will be acyclic. Let $G_1 = (\mathcal{U}_1 \cup \mathcal{V}_1, E_1)$ be the parent graph for model M_1 and let $G_2 = (\mathcal{U}_2 \cup \mathcal{V}_2, E_2)$ be the parent graph for model M_2. Let $G' = ((\mathcal{U}_1 \cup \mathcal{V}_1) \cup (\mathcal{U}_2 \cup \mathcal{V}_2), E_1 \cup E_2)$. In linear time, we can compute whether G' contains any cycles. If it does not, then $M_1 \oplus' M_2$ is guaranteed to be acyclic. This is a sufficient but not necessary condition for acyclicity, as edges can be deleted via our combination process. In practice, though, we suspect this condition will hold in most cases of interest where the combined model is indeed acyclic.

3.3 Combination as least upper bound

When we combine two models, we would like the combined model to be the simplest model that can explain both. Unfortunately, this may not be the case for $M_1 \oplus' M_2$. Indeed, even if M_1 and M_2 are compatible, $M_1 \oplus' M_2$ may not be able to explain both M_i for all variables C that appear in M_i. It follows from Proposition 3.6 that if $M_i \succeq_C M_{3-i}$ or $C \in (\mathcal{U}_i \cup \mathcal{V}_i) - (\mathcal{U}_{3-i} \cup \mathcal{V}_{3-i})$, then $M_1 \oplus' M_2 \equiv_C M_i$, so (by Proposition 3.4) $M_1 \oplus' M_2$ can explain M_i with respect to C. But, as the following example shows, if $M_1 \succeq_C M_2$ and C appears in M_2, $M_1 \oplus' M_2$ may not be able to explain M_2 with respect to C.

Example 3.8. Consider the models M_1 and M_2 depicted in Figure 4, where the range of all variables is $\{\mathsf{true}, \mathsf{false}\}$; the focus set of each variable consists of just its parents, as defined in the parent graph; in M_1, the structural equations are such that $C = A \,\mathsf{XOR}\, B$, while in M_2, $A = D$ and $B = D$. Then in $M_1 \oplus' M_2$, all three of these equations hold.

It is easy to see that $M_1 \oplus' M_2 \not\succeq_C M_2$. The problem is, to explain the value $C = 0$, A and B need to have different values, and there is no context in $M_1 \oplus' M_2$ that gives them different values. Intuitively, although M_2 can explain M_1 wth respect to each of A and B individually, it cannot explain them both together. In particular, the setting $A = \mathsf{false}$ and $B = \mathsf{true}$ cannot be explained in M_2. We have not defined what it would mean to explain a setting involving more than one variable; this is because our intuition for "can explain" is based on the assumption that experts are testing one variable at a time.

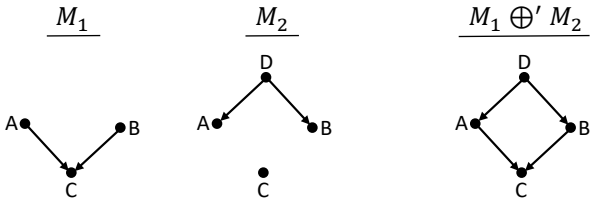

Figure 4: Models M_1 and M_2 where $M_1 \oplus' M_2 \not\succeq_C M_2$.

This example also shows that $\succeq_C$ is not necessarily transitive: we have $M_1 \oplus' M_2 \succeq_C M_1$ and $M_1 \succeq_C M_2$, we do not have $M_1 \oplus' M_2 \succeq_C M_2$. This does not contradict Proposition 3.2, since $M_1 \oplus' M_2$ and M_2 are not compatible. □

The fact that $M_1 \oplus' M_2$ may not be able to explain both M_1 and M_2 is somewhat disconcerting. However, the situation is not quite as bad as it appears.

Definition 3.9. M_1 dominates M_2, *written* $M_1 \succeq M_2$, *if* $M_1 \succeq_C M_2$ *for all* $C \in \mathcal{U}_2 \cup \mathcal{V}_2$.

Note that if M_1 dominates M_2, then we must have that $\mathcal{U}_1 \cup \mathcal{V}_1 \supseteq \mathcal{U}_2 \cup \mathcal{V}_2$.

Theorem 3.10. *If* M_1 *and* M_2 *are compatible, then* $M_1 \oplus' M_2$ *dominates both* M_1 *and* M_2 *iff* $M_1 \oplus' M_2$ *is the unique least upper bound of* $\{M_1, M_2\}$.

Proof. Suppose that $M_1 \oplus' M_2 \succeq M_1$ and $M_1 \oplus' M_2 \succeq M_2$. Then, by definition, $M_1 \oplus' M_2$ is an upper bound of $\{M_1, M_2\}$, so now we must show that, for any other upper bound M' of $\{M_1, M_2\}$, we have $M' \succeq M_1 \oplus' M_2$. We first note that the variables in $M_1 \oplus' M_2$ are precisely $(\mathcal{U}_1 \cup \mathcal{V}_1) \cup (\mathcal{U}_2 \cup \mathcal{V}_2)$. For each variable C in $M_1 \oplus' M_2$, there exists some $i \in \{1, 2\}$ such that $\mathcal{G}^{M_1 \oplus' M_2}(C) = \mathcal{G}_i(C)$ and either $\mathcal{F}^{M_1 \oplus' M_2}(C) = \mathcal{F}_i(C)$ or C is exogenous in both $M_1 \oplus' M_2$ and M_i. Moreover, $Par^{M_1 \oplus' M_2}(C) = Par_i(C)$. Thus, given an intervention $\mathcal{G}^{M_1 \oplus' M_2}(C) = \vec{x}$ in $M_1 \oplus' M_2$ and context $\vec{u}$ such that $(M_1 \oplus' M_2, \vec{u}) \models [\mathcal{G}^{M_1 \oplus' M_2}(C) = \vec{x}](C = c)$, there exists a context $\vec{u}'$ in M_i such that $(M_i, \vec{u}') \models [\mathcal{G}^{M_1 \oplus' M_2}(C) = \vec{x}](C = c)$. Since $M' \succeq M_i$, it follows that there exists a context $\vec{u}''$ in M' such that $(M', \vec{u}'') \models [\mathcal{G}^{M_1 \oplus' M_2}(C) = \vec{x}](C = c)$. It follows that $M' \succeq_C M_1 \oplus' M_2$. Since C was arbitrary, it follows that $M' \succeq M_1 \oplus' M_2$.

We have thus shown that $M_1 \oplus' M_2$ is a least upper bound of $\{M_1, M_2\}$ if $M_1 \oplus' M_2 \succeq M_1$ and $M_1 \oplus' M_2 \succeq M_2$. Uniqueness is straightforward: if M' is another least upper bound of $\{M_1, M_2\}$ then, by Proposition 3.4, it follows that

$M' \equiv_C M_1 \oplus' M_2$ for all $C \in \mathcal{U}_1 \cup \mathcal{V}_1 \cup \mathcal{U}_2 \cup \mathcal{V}_2$, so $M' = M_1 \oplus' M_2$. The converse is also immediate: if $M_1 \oplus M_2$ is not an upper bound of both M_1 and M_2, it certainly cannot be a least upper bound of $\{M_1, M_2\}$. □

So where does this leave us? Our goal is to combine the information of experts. If a decision-maker believes that models M_1 and M_2 both provide useful information, then she would want to work with a model that somehow combines this information. As Example 3.8 shows, the problem with $M_1 \oplus' M_2$ is that it does not necessarily combine all the information in M_1 and M_2. To deal with this problem, we simply define $\oplus$ by taking $M_1 \oplus M_2 = M_1 \oplus' M_2$ if $M_1 \oplus' M_2 \succeq M_i$ for $i = 1, 2$, and otherwise say that M_1 and M_2 are incompatible and $M_1 \oplus M_2$ is undefined. Intuitively, in the latter case, there is no clear way to explain both models, so more experiments are necessary. It is easy to check that Proposition 3.6 holds for $\oplus$, with no change in proof. Moreover, by Proposition 3.10, when it is defined, $M_1 \oplus M_2$ is the least upper bound of $\{M_1, M_2\}$.

We conjecture that if $M_1 \oplus M_2$ is not defined, then $\{M_1, M_2\}$ in fact has no least upper bound. This is the case in the models of Example 3.8. Consider the models M_1' and M_2', where M_1' is identical to M_1 except that it includes the variable D, and $\mathcal{G}^{M_1'}(A) = \mathcal{G}^{M_1'}(B) = \{D\}$, and M_2' is just like M_2 except that $\mathcal{G}^{M_2'}(C) = \{A, B\}$. It is easy to check that M_1 and M_2 are both upper bounds on $\{M_1, M_2\}$, and there is no upper bound M' of $\{M_1, M_2\}$ such that $M_1' \succeq M'$ and $M_2' \succeq M'$.

If this conjecture is correct (and we have shown that it is in a number of special cases), then it shows that if we think of $\succeq$ as an information ordering, then $M_1 \oplus M_2$, when it is defined, is the model that combines the information in M_1 and M_2 and has no additional information; if it is not defined, then there is no such model.[3]

3.4 Explanation complexity and combination complexity

Recall that $M_1 \succeq_C M_2$ if, for every intervention $\mathcal{G}_2(C) = \vec{x}$, value $c \in \mathcal{R}(C)$, and context $\vec{u}_2$, there exists a context $\vec{u}_1$ such that if $(M_2, \vec{u}_2) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$ then $(M_1, \vec{u}_1) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$. However, in principle, we could use a different context $\vec{u}_1$ to explain each possible intervention on $\mathcal{G}_2(C)$. We might be reluctant to accept explanations that are complicated, in the sense of requiring too many different contexts; if an overly complicated explanation is needed to reconcile two models, we may instead prefer to simply declare them incompatible. The following definitions of *explanation complexity* and *combination complexity* capture these intuitions.

Definition 3.11. M_1 can explain M_2 with respect to C using a set $\mathcal{U}'$ of contexts *if M_1 can explain M_2 with respect to C using only contexts u_1' drawn from $\mathcal{U}_1$; that is, we just modify Definition 3.1 so that all the contexts u_1 in condition (c) are drawn from $\mathcal{U}'$. The* complexity of M_1's ability to explain M_2 with respect to C *is* $\min\{|\mathcal{U}'| : M_1$ *can explain* M_2 *with respect to* C *using* $\mathcal{U}'\}$.

Example 3.12. Consider the models in Figure 5. In all of these models, $\mathcal{R}(A) = \mathcal{R}(B) = \mathcal{R}(A_1) = \mathcal{R}(A_2) = \mathcal{R}(A_3) = \{0, \ldots, 10\}$, $\mathcal{R}(D) = \{0, \ldots, 30\}$, and $\mathcal{R}(C) = \{0, \ldots, 60\}$. In model M_1 on the left, we have the structural equations $C = A + B$ if $D \geq 1$ and $C = 2(A + B)$ if $D = 0$; in model M_2, we have $C = A + B$; in model M_3 on the right, we have $C = D$; and in model M_4, we have $C = A_1 + A_2 + A_3$. In the low-complexity models on the left, the complexity of M_1's ability to explain M_2 with respect to C is 1, as every intervention can be explained by the value of D simply having been 1 the entire time. For the high-complexity models on the right, though, the complexity of M_3's ability to explain M_4 with respect to C is 30; for each intervention, D must take on precisely the right value in M_3 for each particular outcome of C to be observed. Thus, we would be more hesitant to combine the high-complexity models M_3 and M_4. Combining them implicitly assumes that M_3 and M_4 are compatible, and, in particular, that M_3 can explain M_4 with respect to C. □

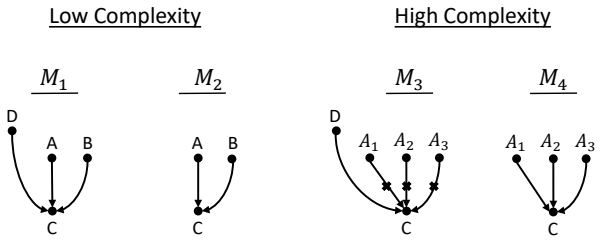

Figure 5: The two models on the left have low explanation complexity with respect to C whereas the two on the right have high explanation complexity.

We can extend the notion of explanation complexity to the combination complexity of two models.

Definition 3.13. *The* combination complexity of two compatible models M_1 and M_2 *is the minimum cardinality* $|\mathcal{U}'|$ *taken over all sets* $\mathcal{U}'$ *such that, for all* $C \in (\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2)$, *either* M_1 *can explain* M_2 *with respect to* C *using* $\mathcal{U}'$ *or* M_2 *can explain* M_1 *with respect to* C *using* $\mathcal{U}'$.

A decision-maker may want to consider only explanations that have complexity less than or equal to some threshold or model combinations that have complexity less than a threshold. In the next section, we show how combination complexity can be used to weight models.

4 Weighting and Combining Expert Opinions

Given a collection of models, it may be impossible to combine all of them, but possible to combine a variety of different subsets of them. ACH proposed a way to assign confidence to different possible combined models based on the

[3] We remark that we can define an analogue of $\succeq$ for the notion of combination considered by ACH, and show that $M_1 \oplus M_2$ as ACH define it is the least upper bound M_1 and M_2 with respect to the ACH notion. Thus, thinking in terms of least upper bound seems like a useful way to think of combining models.

decision-maker's confidence in the original models. Here we provide a way to extend this to our setting.

We start with a collection of pairs $(M_1, p_1), \ldots, (M_n, p_n)$ where M_i is a causal model with focus and p_i is a value in $(0, 1]$. Here the intuition for each pair should be that M_i was the model proposed by expert i and p_i is the decision-maker's degree of confidence that expert i's model is correct. More precisely, p_i is not the decision-maker's degree of confidence that the assumptions built into M_i are correct, but her confidence that, for each variable C and intervention $\mathcal{G}_i(C) = \vec{x}$, if $(M_i, \vec{u}) \models [\mathcal{G}_i(C) = \vec{x}](C = c)$ then expert i indeed observed a world where the variables in $\mathcal{G}_i(C)$ were $\vec{x}$ and C did have value c. Following ACH, we define $Compat = \{I \subseteq \{1, \ldots, n\} :$ the models in $\{M_i : i \in I\}$ are mutually compatible$\}$ and define $M_I = \oplus_{i \in I} M_i$ for all $I \in Compat$. The *mutual compatibility* of a set $\mathcal{M}$ of models is defined inductively on the cardinality of $\mathcal{M}$. If $|\mathcal{M}| = 1$ then $\mathcal{M}$ is automatically mutually compatible, and if $|\mathcal{M}| = 2$ then $\mathcal{M}$ is mutually compatible if the two models in $\mathcal{M}$ are compatible. If $|\mathcal{M}| = n$ then $\mathcal{M}$ is mutually compatible if every subset of cardinality $n - 1$ is mutually compatible and, for each $M \in \mathcal{M}$, M is compatible with $\oplus_{M' \in \mathcal{M}: M' \neq M} M'$.

One simple way to weight the combined models, proposed by ACH, is to assign model M_I probability

$$p_I = \prod_{i \in I} p_i * \prod_{j \notin I} (1 - p_j)/N, \tag{1}$$

where N is simply a normalization term to get the probabilities to sum to 1. Thus, p_I captures the intuition that the agents in I performed their experiments correctly while the agents not in I may have made a mistake in one or more of their experiments, where the probabilities of agents having made a mistake are treated as being mutually independent.

Let $\mathcal{M}_I = \{M_i : i \in I\}$. In our setting, we may also want to take into account how complex it is to combine the models in $\mathcal{M}_I$ when assigning M_I a probability; if combining the models in $\mathcal{M}_I$ requires a large set of contexts to make all of the necessarily explanations, then we may have less confidence that the combined model captures the true state of the world. To formalize this idea, we first generalize Definition 3.13 in the obvious way: the combination complexity of a set $\mathcal{M}$ is the minimum cardinality $|\mathcal{U}'|$ of a set $\mathcal{U}'$ such that all explanations made during the combination process can be made using $\mathcal{U}'$. The combination complexity of a singleton set is defined to be 1.

Exactly how complexity should be taken into account when assigning confidence scores may be context-dependent; it is up to the decision-maker who is combining the models to decide. We propose several simple rules here. One simple rule that may be relevant in some situations is to simply use a threshold, and assign confidence 0 to models where the combination complexity or the explanation complexity with respect to any variable C is above some constant μ. (Here and in the following two rules, the normalization factor N must be updated accordingly.) Another natural option may be to add a weighting factor to (1) that is inversely proportional to the combination complexity. If the combination complexity of $\mathcal{M}_I$ is μ_I, then we set

$$p'_I = \frac{1}{\mu_I} * \prod_{i \in I} p_i * \prod_{j \notin I} (1 - p_j)/N.$$

A third rule that may be useful in some contexts is to assign complexity weights that are inverse exponential in the combination complexity. Here the confidence scores assigned would be

$$p''_I = e^{-\mu_I} * \prod_{i \in I} p_i * \prod_{j \notin I} (1 - p_j)/N.$$

Example 4.1. Consider three models M_1, M_2, and M_3, where

- $\mathcal{U}_1 = \mathcal{U}_3 = \{A, B, D\}$, $\mathcal{V}_1 = \mathcal{V}_3 = \{C\}$, $\mathcal{U}_2 = \{A, G\}$, and $\mathcal{V}_2 = \{B, C\}$;
- $\mathcal{R}_1(C) = \mathcal{R}_2(C) = \mathcal{R}_3(C) = \mathcal{R}_1(D) = \mathcal{R}_3(D) = \{0, 1, 2\}$ and $\mathcal{R}_1(A) = \mathcal{R}_1(B) = \mathcal{R}_2(A) = \mathcal{R}_2(B) = \mathcal{R}_2(G) = \mathcal{R}_3(A) = \mathcal{R}_3(B) = \{0, 1\}$;
- $\mathcal{G}_1(C) = \{A, B, D\}$, $\mathcal{G}_2(C) = \{A, B\}$, $\mathcal{G}_2(B) = \{G\}$, and $\mathcal{G}_3(C) = \{A, B, D\}$;
- the structural equations are such that, in M_1, $C = D$; in M_2, $C = A + B$ and $B = G$; and in M_3, $C = 2$ if $D = 0$ and $C = \min(1, A + B)$ if $D = 1$ or $D = 2$.

The models in the set $\{M_I : I \in Compat\}$ are M_1, M_2, M_3, $M_1 \oplus M_2$, and $M_3 \oplus M_2$, with combination complexity 5 for $M_1 \oplus M_2$ (3 for M_1 to explain M_2 with respect to C and 2 for M_2 to explain M_1 with respect to B) and combination complexity 4 for $M_3 \oplus M_2$ (2 for M_3 to explain M_2 with respect to C and 2 for M_2 to explain M_3 with respect to B). Of course, M_1, M_2, and M_3 (viewed as singleton sets) all have combination complexity 1, by definition. Consider the second weighting rule above, inversely proportional weighting, with prior confidences $p_1 = 0.85, p_2 = 0.8,$ and $p_3 = 0.9$. The assigned confidence scores would then be

$$\begin{aligned}
p'_{M_1} &= (0.85)(0.2)(0.1)/N \approx 0.176 \\
p'_{M_2} &= (0.15)(0.8)(0.1)/N \approx 0.124 \\
p'_{M_3} &= (0.15)(0.2)(0.9)/N \approx 0.280 \\
p'_{M_1 \oplus M_2} &= (\tfrac{1}{5})(0.85)(0.8)(0.1)/N \approx 0.141 \\
p'_{M_3 \oplus M_2} &= (\tfrac{1}{4})(0.15)(0.8)(0.9)/N \approx 0.280.
\end{aligned}$$

Under the third rule, inverse exponential weighting, with the same prior confidences, the assigned confidence scores would be

$$\begin{aligned}
p''_{M_1} &= (0.85)(0.2)(0.1)/N \approx 0.291 \\
p''_{M_2} &= (0.15)(0.8)(0.1)/N \approx 0.205 \\
p''_{M_3} &= (0.15)(0.2)(0.9)/N \approx 0.462 \\
p''_{M_1 \oplus M_2} &= (e^{-5})(0.85)(0.8)(0.1)/N \approx 0.008 \\
p''_{M_3 \oplus M_2} &= (e^{-4})(0.15)(0.8)(0.9)/N \approx 0.034.
\end{aligned}$$

As expected, the inverse exponential weighting rule is more complexity averse, and so assigns a greater proportion of confidence to the uncombined models. □

These three rules behave in a qualitatively similar manner, with the importance of complexity being taken into account in different ways. More generally, let μ_I be the combination

complexity of $\mathcal{M}_I$ and let $Q_I = \prod_{i \in I} p_i * \prod_{j \notin I}(1 - p_j)$. We believe that there are many reasonable functions $f(Q_I, \mu_I)$ that can be used to assign a confidence scores to M_I; we leave it up to the decision-maker to decide what function f is most suitable for a given context. The two requirements that seem necessary to us is that f be non-increasing in μ_I and non-decreasing in Q_I; that is, $f(Q_I, \mu_I) \geq f(Q_I, \mu'_I)$ for fixed Q_I if $\mu'_I \geq \mu_I$, and $f(Q_I, \mu_I) \leq f(Q'_I, \mu_I)$ for fixed μ_I if $Q'_I \geq Q_I$. These two rules capture the intuition that we should not prefer models that are more complicated, nor should we prefer models that are composed of models in which we had less prior confidence.

An additional factor that may sometimes play a role is the likelihood of different endogenous settings occurring. If one model can explain the other only by using a context that is very unlikely to occur, then we may not want to assign much weight to that combined model. Thus, in certain settings it may also make sense to have the confidence scores depend on a distribution over exogenous settings.

5 Computational Complexity

We now consider the computational complexity of determining whether one model can explain another with respect to C.

Theorem 5.1. *Determining whether $M_1 \succeq_C M_2$ is in Π_2^P, and is Π_2^P-hard, even in instances where all variables are binary.*

Proof. It is easy to see that the problem is in Π_2^P: the first two conditions in the can-explain relation can clearly be checked in polynomial time, while, for a fixed intervention $\mathcal{G}_2(C) = \vec{x}$ in M_2, context $\vec{u}_2$ in M_2, and context $\vec{u}_1$ in M_1, checking whether $(M_2, \vec{u}_2) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$ and $(M_1, \vec{u}_1) \models [\mathcal{G}_2(C) \leftarrow \vec{x}](C = c)$ can be done in polynomial time.

For the lower bound, consider the canonical Π_2^P-hard language Π_2^P(SAT)=$\{\forall\vec{X}\exists\vec{Y}\varphi$:$\forall\vec{X}\exists\vec{Y}\varphi$ is a closed quantified Boolean formula, $\forall\vec{X}\exists\vec{Y}\varphi =$ **true**$\}$. We show a reduction from Π_2^P(SAT) to our language.

Consider a CQBF (closed quantified Boolean formula) $\forall\vec{X}\exists\vec{Y}\varphi$; we show how to transform this into an instance of our problem. For ease of exposition, we assume that all variables in $\vec{X} \cup \vec{Y}$ appear in φ. Let M_2 contain exogenous variables $\vec{X}$ and an endogenous variable $C \notin \vec{X} \cup \vec{Y}$. In M_2, the range of all variables is $\{\mathsf{true}, \mathsf{false}\}$, $\mathcal{G}_2(C) = \vec{X}$, and the equation for C is $C = \mathsf{true}$. In M_1, we have $\mathcal{U}_1 = \vec{X} \cup \vec{Y}$, $\mathcal{V}_1 = \{C\}$, $\mathcal{G}_1(C) = \vec{X} \cup \vec{Y}$, and the equation for C is $C = \varphi$.

We now show that $M_1 \succeq_C M_2$ if and only $\forall\vec{X}\exists\vec{Y}\varphi$ is true. First, suppose that $M_1 \succeq_C M_2$. Because C is always true in M_2 and $\mathcal{G}_2(C) = \vec{X}$, by condition (c) in the definition of the can-explain relation, for all settings of $\vec{X}$ there must be a setting of the remaining variables such that $C = \mathsf{true}$ in M_1. But because the equation for C in M_1 is $C = \varphi$, this means that for all settings of $\vec{X}$, there exists a setting of $\vec{Y}$ such that φ is true. For the other direction, suppose that $\forall\vec{X}\exists\vec{Y}\varphi$ is true. Clearly $\mathcal{G}_2(C) \subseteq \mathcal{G}_1(C)$ and $\mathcal{R}_1(C) = \mathcal{R}_2(C)$. To see that condition (c) of the definition of can-explain holds, consider an intervention $\vec{X} = \vec{x}$ on $\vec{X}$. Because $\forall\vec{X}\exists\vec{Y}\varphi$ is true, there must be some setting of the values in $\vec{Y}$ such that if $\vec{X}$ were set to $\vec{x}$, then C would evaluate to true in M_1. So in the context where $\vec{Y}$ is set correspondingly, we get that the original intervention would make $C = \mathsf{true}$, as desired. □

While this result indicates that this computation is likely to be intractable in the worst case, models that arise in the physical and social sciences often contain only a small number of variables, so we would still expect these computations to be feasible in practice.

6 Conclusion

We have shown how causal models can be combined in instances where experts disagree due to different focus areas. We defined what it means for one model to be able to explain another with respect to a given variable and showed how this can be used to combine two compatible models. Furthermore, we showed that the model obtained via this combination process is in fact the least upper bound of the combined models relative to the natural relation, in some sense making it the simplest model that can explain the observations of both experts.

The can-explain relation embodies one way of explaining why two experts may have different causal models. ACH can be viewed as modeling a different reason, where M_1 is "better than" M_2 with respect to a variable C in the ACH view if, roughly speaking, M_1 has a more detailed picture of the causal relations among the ancestors of C. While we believe that the can-explain relation captures quite a natural intuition (as does the ACH notion of compatibility!), there may well be other reasonable intuitions that are worth exploring. More generally, this viewpoint suggests that a decision-maker trying to combine experts' models must think seriously about the reasons underlying their disagreement before combining models, and consider a notion of combination appropriate for these reasons. Since the need to combine expert opinions arises frequently in practice, having a principled understanding of the process seems to us critical. We hope that the results of this paper help in this process.

Acknowledgments: This work 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.

References

Alrajeh, D.; Chockler, H.; and Halpern, J. Y. 2018. Combining experts' causal judgments. In *Proc. Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*.

Bradley, R.; Dietrich, F.; and List, C. 2014. Aggregating causal judgments. *Philosophy of Science* 81(4):419–515.

Claassen, T., and Heskes, T. 2010. Learning causal network structure from multiple (in)dependence models. In *Proc. of the Fifth European Workshop on Probabilistic Graphical Models*, 81–88.

Claassen, T., and Heskes, T. 2012. A Bayesian approach to constraint based causal inference. In *Proc. 28th Conference on Uncertainty in Artificial Intelligence (UAI 2012)*, 207–217.

Halpern, J. Y., and Pearl, J. 2001. Causes and explanations: A structural-model approach — Part II: Explanation. In *Proc. Seventeenth International Joint Conference on Artificial Intelligence (IJCAI '01)*, 27–34.

Hyttinen, A.; Eberhardt, F.; and Jarvisalo, M. 2014. Constraint-based causal discovery: conflict resolution with answer set programming. In *Proceedings of the Thirtieth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-14)*, 340–349. Corvallis, Oregon: AUAI Press.

Tillman, R. E., and Spirtes, P. 2011. Learning equivalence classes of acyclic models with latent and selection variables from multiple datasets with overlapping variables. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2011, Fort Lauderdale, USA, April 11-13, 2011*, volume 15 of *JMLR Proceedings*, 3–15. JMLR.org.

Triantafillou, S., and Tsamardinos, I. 2015. Constraint-based causal discovery from multiple interventions over overlapping variable sets. *Journal of Machine Learning Research* 16:2147–2205.