Combining Experts' Causal Judgments

Dalal Alrajeh
Department of Computing
Imperial College London
dalal.alrajeh04@imperial.ac.uk

Hana Chockler
Department of Informatics
King's College London
Hana.Chockler@kcl.ac.uk

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

Abstract

Consider a policymaker who wants to decide which intervention to perform in order to change a currently undesirable situation. The policymaker has at her disposal a team of experts, each with their own understanding of the causal dependencies between different factors contributing to the outcome. The policymaker has varying degrees of confidence in the experts' opinions. She wants to combine their opinions in order to decide on the most effective intervention. We formally define the notion of an effective intervention, and then consider how experts' causal judgments can be combined in order to determine the most effective intervention. We define a notion of two causal models being *compatible*, and show how compatible causal models can be combined. We then use it as the basis for combining experts causal judgments. We illustrate our approach on a number of real-life examples.

1 Introduction

Consider a policymaker who is trying to decide which intervention, that is, which actions, should be implemented in order to bring about a desired outcome, such as preventing violent behavior in prisons or reducing famine mortality in some country. The policymaker has access to various experts who can advise her on which interventions to consider. Some experts may be (in the policymaker's view) more reliable than others; they may also have different areas of expertise; or may have perceived alternative factors in their analysis. The goal of the policymaker is to choose the best intervention, taking into account the experts' advice.

There has been a great deal of work on combining experts' probabilistic judgments. (Genest and Zidek (1986) provide a somewhat dated but still useful overview; Dawid (1987) and Fenton et al. (2016), among others, give a Bayesian analysis.) We are interested in combining experts' judgments in order to decide on the best intervention. Hence, we need more than probabilities. We need to have a causal understanding of the situation. Thus, we assume that the experts provide the policymaker with *causal models*. In general, these models may involve different variables (since the experts may be focusing on different aspects of the problem). Even if two models both include variables C_1 and C_2, they may disagree on the relationships between them. For example, one expert may believe that C_2 is independent of C_1 while another may believe that C_1 causally depends on C_2. Yet somehow the policymaker wants to use the information in these causal models to reach her decision.

Despite the clear need for causal reasoning, and the examples in the literature and in practice where experts work with causal models (e.g., (Chockler et al. 2015; Sampson, Winship, and Knight 2013)), there is surprisingly little work on combining causal judgments. Indeed, the only work that we are aware of is that of Bradley, Dietrich, and List (2014) (BDL from now on), who prove an impossibility result. Specifically, they describe certain arguably reasonable desiderata, and show that there is no way of combining causal models so as to satisfy all their desiderata. They then discuss various weakenings of their assumptions to see the extent to which the impossibility can be avoided, none of which seem that satisfactory.

There is also much work on the closely related problem of *causal discovery*: constructing a single causal model from a data set. A variety of techniques have been used to find the model that best describes how the data is generated (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). Of course, if we have the data that the experts used to generate their models, then we should apply the more refined techniques of the work on causal discovery. However, while the causals model constructed by experts are presumably based on data, the data itself is typically no longer available. Rather, the models represent the distillation of years of experience, obtained by querying the experts.

In this paper, we present an approach to combining experts' causal models when sufficient data for discovering the overall causal model is not available. The key step in combining experts' causal models lies in defining when two causal models are *compatible*. Causal models can be combined only if they are compatible. We start with a notion of *strong* compatibility, where the conditions say, among other things, that if both M_1 and M_2 involve variables C_1 and C_2, then they must agree on the causal relationship between C_1 and C_2. But that is not enough. Suppose that in both models C_1 depends on C_2, C_3, and C_4. Then in a precise sense, the two models must agree on *how* the dependence works, despite describing the world using possibly different sets of variables. Roughly speaking, this is the case when, for every

variable C that the two models have in common, we can designate one of the models as being "dominant" with respect to C, and use that model to determine the relationships for C. When M_1 and M_2 are compatible, we are able to construct a combined model $M_1 \oplus M_2$ that can be viewed as satisfying all but one of BDL's desiderata (and we argue that the one it does not satisfy is unreasonable).

This set of constraints is very restrictive, and, as we show on real-life examples, models are often not compatible in this strong sense. We thus define two successively more general notions of compatibility. But even with this more general notions, we may find that not all the experts' models are incompatible. In that case, we simply place a probability on possible ways of combining the compatible models, using relatively standard techniques, based on the perceived reliability of the experts who proposed them. The policymaker will then have a probability on causal models that she can use to decide which interventions to implement. Specifically, we can use the probability on causal models to compute the probability that an intervention is efficacious. Combining that with the cost of implementing the intervention, she can compute the most effective intervention. As we shall see, although we work with the same causal structures used to define causality, interventions are different from (and actually simpler to analyze than) causes.

We believe that our approach provides a useful formal framework that can be applied to the determination of appropriate interventions in real-world scenarios involving complex sociological phenomena, such as crime prevention scenarios (Sampson, Winship, and Knight 2013) and radicalization (Wikström and Bouhana 2017).

The rest of the paper is organized as follows. Section 2 provides some background material on causal models. We formally define our notion of intervention and compare it to causality in Section 3. We discuss our concept of compatibility and how causal models can be combined in Section 4. We discuss how the notions of interventions and of compatible models can be used by the policymakers to choose optimal interventions in Section 5. Finally, in Section 6, we illustrate these concepts on two case studies before concluding in Section 7.

2 Causal Models

In this section, we review the definition of causal models introduced by Halpern and Pearl (2005). The material in this section is largely taken from (Halpern 2016).

We assume that the world is described in terms of variables and their values. Some variables may have a causal influence on others. This influence is modeled by a set of *structural equations*. It is conceptually useful to split the variables into two sets: the *exogenous* variables, whose values are determined by factors outside the model, and the *endogenous* variables, whose values are ultimately determined by the exogenous variables. For example, in a voting scenario, we could have endogenous variables that describe what the voters actually do (i.e., which candidate they vote for), exogenous variables that describe the factors that determine how the voters vote, and a variable describing the outcome (who wins). The structural equations describe how these values are determined (e.g., majority rules; a candidate wins if A and at least two of B, C, D, and E vote for him; etc.).

Formally, a *causal model* M is a pair $(\S, \mathcal{F})$, where $\S$ is a *signature*, which explicitly lists the endogenous and exogenous variables and characterizes their possible values, and $\mathcal{F}$ defines a set of *(modifiable) structural equations*, relating the values of the variables. A signature $\S$ 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 $\mathcal{R}$ associates with every variable $Y \in \mathcal{U} \cup \mathcal{V}$ a nonempty set $\mathcal{R}(Y)$ of possible values for Y (that is, the set of values over which Y *ranges*). For simplicity, we assume here that $\mathcal{V}$ is finite, as is $\mathcal{R}(Y)$ for every endogenous variable $Y \in \mathcal{V}$. $\mathcal{F}$ associates with each endogenous variable $X \in \mathcal{V}$ a function denoted F_X (i.e., $F_X = \mathcal{F}(X)$) such that $F_X : (\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V} - \{X\}} \mathcal{R}(Y)) \rightarrow \mathcal{R}(X)$. This mathematical notation just makes precise the fact that F_X determines the value of X, given the values of all the other variables in $\mathcal{U} \cup \mathcal{V}$. If there is one exogenous variable U and three endogenous variables, X, Y, and Z, then F_X defines the values of X in terms of the values of Y, Z, and U. 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.[1]

The structural equations define what happens in the presence of external interventions. Setting the value of some variable X to x in a causal model $M = (\S, \mathcal{F})$ results in a new causal model, denoted $M_{X \leftarrow x}$, which is identical to M, except that the equation for X in $\mathcal{F}$ is replaced by $X = x$.

The dependencies between variables in a causal model M can be described using a *causal network* (or *causal graph*), whose nodes are labeled by the endogenous and exogenous variables in $M = ((\mathcal{U}, \mathcal{V}, \mathcal{R}), \mathcal{F})$, with one node for each variable in $\mathcal{U} \cup \mathcal{V}$. The roots of the graph are (labeled by) the exogenous variables. There is a directed edge from variable X to Y if Y *depends on* X; this is the case if there is some setting of all the variables in $\mathcal{U} \cup \mathcal{V}$ other than X and Y such that varying the value of X in that setting results in a variation in the value of Y; that is, there is a setting $\vec{z}$ of the variables other than X and Y and values x and x' of X such that $F_Y(x, \vec{z}) \neq F_Y(x', \vec{z})$. A causal model M is *recursive* (or *acyclic*) if its causal graph is acyclic. It should be clear that if M is an acyclic causal model, then given a *context*, that is, a setting $\vec{u}$ for the exogenous variables in $\mathcal{U}$, the values of all the other variables are determined (i.e., there is a unique solution to all the equations). We can determine these values by starting at the top of the graph and working our way down. In this paper, following the literature, we restrict to recursive models.

In many papers in the literature (e.g., (Bradley, Dietrich, and List 2014; Sampson, Winship, and Knight 2013)) a causal model is defined simply by a causal graph indicating the dependencies, perhaps with an indication of whether whether a change has a positive or negative effect; that is, edges are annotated with $+$ or $-$, so that an edge from A to

[1] The fact that X is assigned $U + Y$ (i.e., the value of X is the sum of the values of U and Y) does not imply that Y is assigned $X - U$; that is, $F_Y(U, X, Z) = X - U$ does not necessarily hold.

B annotated with $+$ means that an increase in A results in an increase in B, while if it is annotated with a $-$, then an increase in A results in a decrease in B (where what constitutes an increase or decrease is determined by the model). Our models are more expressive, since the equations typically provide much more detailed information regarding the dependency between variables; the causal graphs capture only part of this information. Of course, this extra information makes combining models more difficult (although, as the results of BDL show, the difficulties in combining models already arise with purely qualitative graphs).

To define interventions carefully, it is useful to have a language in which we can make statements about 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 §)* 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 $i = 1, \ldots, k$.

We call a pair $(M, \vec{u})$ consisting of a causal model M and a context $\vec{u}$ a *(causal) setting*. A causal formula ψ is true or false in a setting. We write $(M, \vec{u}) \models \psi$ if the causal formula ψ is true in the setting $(M, \vec{u})$. 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 acyclic models) solution to the equations in M in context $\vec{u}$ (that is, the unique vector of values for the exogenous variables that simultaneously satisfies all equations in M with the variables in $\mathcal{U}$ set to $\vec{u}$). Finally, $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\varphi$ if $(M_{\vec{Y}=\vec{y}}, \vec{u}) \models \varphi$.

3 Interventions

In this section we define (causal) interventions, and compare the notion of intervention to that of cause.

Definition 1 *[Intervention]* $\vec{X} = \vec{x}$ *is an* intervention on φ in $(M, \vec{u})$ *if the following three conditions hold:*

I1. $(M, \vec{u}) \models \varphi$.

I2. $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}]\neg\varphi$.

I3. $\vec{X}$ *is minimal; there is no strict subset* $\vec{X}'$ *of* $\vec{X}$ *and values* $\vec{x}'$ *such that* $\vec{X}' = \vec{x}'$ *satisfies I2.*

I1 says φ must be true in the current setting $(M, \vec{u})$, while I2 says that performing the intervention results in φ no longer being true. I3 is a minimality condition. From a policymaker's perspective, I2 is the key condition. It says that by making the appropriate changes, we can bring about a change in φ.

Our definition of intervention slightly generalizes others in the literature. Pearl (2009) assumes that the causal model is first analyzed, and then a new intervention variable I_V is added for each variable V on which we want to intervene. If $I_V = 1$, then the appropriate intervention on V takes place, independent of the values of the other parents of V; if $I_V = 0$, then I_V has no effect, and the behavior of V is determined by its parents, just as it was in the original model. Lu and Druzdzel (2002), Korb et al. (2004), and Woodward (2003) take similar approaches.

We do not require special intervention variables; we just allow interventions directly on the variables in the model. But we can certainly assume as a special case that for each variable V in the model there is a special intervention variable I_V that works just like Pearl's intervention variables, and thus recover the other approaches considered in the literature. In any case, it should be clear that all these definitions are trying to capture exactly the same intuitions, and differ only in minor ways.

Although there seems to be relatively little disagreement about how to capture intervention, the same cannot be said for causality. Even among definitions that involve counterfactuals and structural equations (Glymour and Wimberly 2007; Halpern 2015; Halpern and Pearl 2005; Hitchcock 2001; 2007; Woodward 2003), there are a number of subtle variations. Fortunately for us, the definition of intervention does not depend on how causality is defined. While we do not get into the details of causality here, it is instructive to compare the definitions of causality and intervention.

For definiteness, we focus on the definition of causality given by Halpern (2015). It has conditions AC1–3 that are analogous of I1–3. Specifically, AC1 says $\vec{X} = \vec{x}$ is a cause of φ in $(M, \vec{u})$ if $(M, \vec{u}) \models (\vec{X} = \vec{x}) \wedge \varphi$ and AC3 is a minimality condition. AC2 is a more complicated condition; it says that there exist values $\vec{x}'$ for the variables in $\vec{X}$, a (possibly empty) subset $\vec{W}$ of variables, and values $\vec{w}$ for the variables in $\vec{W}$ such that $(M, \vec{u}) \models \vec{W} = \vec{w}$ and $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w}]\neg\varphi$. We do not attempt to explain or motivate AC2 here, since our focus is not causality. The following example, due to Lewis (2000), illustrates some of the subtleties, and highlights the differences between causality and intervention.

Suppose that Suzy and Billy both pick up rocks and throw them at a bottle. Suzy's rock gets there first, shattering the bottle. Since both throws are perfectly accurate, Billy's would have shattered the bottle had Suzy not thrown. Most people would say that Suzy is a cause, and not Billy. Part of the difficulty in getting a good definition of causality is to ensure that the definition gives us this result (given an appropriate causal model). However, Suzy's throw by itself is not an intervention for the bottle shattering. Even if we prevent Suzy from throwing, the bottle will still shatter because of Billy's throw. That is, if we have variables *ST* and *BT* for Suzy's throw and Billy's throw, with possible values 0 and 1 ($ST = 1$ if Suzy throws, $ST = 0$ if she doesn't, and similarly for Billy), then although $ST = 1$ is a cause of the bottle shattering, $ST = 0$ is not an intervention for the bottle shattering; intervening on *ST* alone does not change the outcome. On the other hand, $ST = 0 \wedge BT = 0$ is an intervention for the bottle shattering, but $ST = 1 \wedge BT = 1$ is not a cause of the bottle shattering; it violates the minimality condition AC3.

It is almost immediate from the definitions that we have the following relationship between interventions and causes:

Proposition 3.1 *If* $\vec{X} = \vec{x}$ *is an intervention for* φ *in* $(M, \vec{u})$

then there is some subset of $\vec{X}'$ of $\vec{X}$ such that $\vec{X}' = \vec{x}'$ is a cause of φ in $(M, \vec{u})$, where $\vec{x}'$ is such that $(M, \vec{u}) \models \vec{X}' = \vec{x}'$. Conversely, if $\vec{X} = \vec{x}$ is a cause of φ in $(M, \vec{u})$ then there is a superset $\vec{X}'$ of $\vec{X}$ and values $\vec{x}'$ such that $\vec{X}' = \vec{x}'$ is an intervention for φ.

Halpern (2015) proved that (for his latest definition) the complexity of determining whether $\vec{X} = \vec{x}$ is a cause of φ in $(M, \vec{u})$ is *DP*-complete, where *DP* consists of those languages L for which there exist a language L_1 in NP and a language L_2 in co-NP such that $L = L_1 \cap L_2$ (Papadimitriou and Yannakakis 1982). It is well known that *DP* is at least as hard as NP and co-NP (and most likely strictly harder). The following theorem shows that the problem of determining whether $\vec{X} = \vec{x}$ is an intervention is in a lower complexity class.

Theorem 3.2 *Given a causal model M, a context $\vec{u}$, and a Boolean formula φ, the problem of determining whether $\vec{X} = \vec{x}$ is an intervention for φ in $(M, \vec{u})$ is co-NP-complete.*

Proof: First, we prove membership in co-NP. It is easy to see that checking conditions I1 and I2 of Definition 1 can be done in polynomial time by simply evaluating φ first in $(M, \vec{u})$ and then in the modified context where the values of $\vec{X}$ are set to $\vec{x}$. Checking whether I3 holds is in co-NP, because the complementary condition is in NP; indeed, we simply have to guess a subset $\vec{X}'$ of $\vec{X}$ and values $\vec{x}'$ and verify that I1 and I2 hold for $\vec{X}' = \vec{x}'$ and φ, which, as we observed, can be done in polynomial time.

For co-NP-hardness, we provide a reduction from UNSAT, which is the language of all unsatisfiable Boolean formulas, to the intervention problem. Given a formula ψ that mentions the set $\vec{V}$ of variables, we construct a causal model M_ψ, context $\vec{u}$, and formula φ such that $\vec{V} = 1$ is an intervention for φ in $(M, \vec{u})$ iff ψ is unsatisfiable.

The set of endogenous variables in M is $\vec{V} \cup \{V', Y\}$, where V' and Y are fresh variables not in $\vec{V}$. Let $\vec{W} = \vec{V} \cup \{V'\}$. There is a single exogenous variable U that determines the value of the variables in $\vec{W}$: we have the equation $V = U$ for each variable $V \in \vec{W}$. The equation for Y is $Y = \vee_{V \in \vec{W}} (V = 0)$ (so $Y = 1$ if at least one variable in $\vec{W}$ is 0). Let φ be $\neg\psi \wedge (Y = 1)$. We claim that $\vec{W} = \vec{1}$ is an intervention for φ in $(M_\psi, 0)$ iff $\psi \in$ UNSAT.

Suppose that $\psi \in$ UNSAT. Then, it is easy to see that $(M, 0) \models \varphi$ (since $\neg\psi$ is valid) and $(M, 0) \models [\vec{W} \leftarrow \vec{1}]\neg\varphi$ (since $(M, 0) \models [\vec{W} \leftarrow \vec{1}](Y = 0)$). To see that I3 holds, suppose by way of contradiction that $\vec{W}' \leftarrow \vec{w}'$ satisfies I1 and I2 for some strict subset $\vec{W}'$ of $\vec{W}$. In particular, we must have $(M, 0) \models [\vec{W}' \leftarrow \vec{w}']\neg\varphi$. We clearly have $(M, 0) \models [\vec{W}' \leftarrow \vec{w}'](Y = 1)$, so we must have $(M, 0) \models [\vec{W}' \leftarrow \vec{w}']\psi$, contradicting the assumption that $\psi \in$ UNSAT. Thus, $\vec{W} \leftarrow \vec{1}$ is an intervention for φ, as desired.

For the converse, suppose that $\vec{W} \leftarrow \vec{1}$ is an intervention for φ. Then we must have $(M, 0) \models [\vec{W}' \leftarrow \vec{w}']\neg\psi$ for all strict subsets $\vec{W}'$ of $\vec{W}$ and all settings $\vec{w}'$ of the variables in $\vec{W}'$. Since, in particular, this is true for all subsets $\vec{W}'$ of $\vec{W}$ that do not involve V', it follows that $\neg\psi$ is true for all truth assignments, so $\psi \in$ UNSAT. ∎

In practice, however, we rarely expect to face the co-NP complexity. For reasons of cost or practicality, we would expect a policymaker to consider interventions on at most k variables, for some small k. The straightforward algorithm that, for a given k, checks all sets of variables of the model M of size at most k runs in time $O(|M|^k)$.

4 Combining Compatible Causal Models

This section presents our definition for compatibility of expert opinions. We consider each expert's opinion to be represented by a causal model and, for simplicity, that each expert expresses her opinion with certainty. (We can easily extend our approach to allow the experts to have some uncertainty about the correct model; see the end of Section 5.) We start with a strong notion of compatibility, and then consider generalizations of this notion that are more widely applicable.

4.1 Domination and Compatibility

To specify what it means for a set of models to be compatible, we first define what it means for the causal model M_1 to contain at least as much information about variable C as the causal model M_2, denoted $M_1 \succeq_C M_2$. Intuitively, M_1 contains at least as much information about C as M_2 if M_1 and M_2 say the same things about the causal structure of C as far as the variables that M_1 and M_2 share, but M_1 contains (possibly) more detailed information about C, because, for example, there are additional variables in M_1 that affect C. We capture this property formally below. Say that B is an *immediate M_2-ancestor of Y in M_1* if $B \in \mathcal{U}_2 \cup \mathcal{V}_2$, B is an ancestor of Y in M_1, and there is a path from B to Y in M_1 that has no nodes in $\mathcal{U}_2 \cup \mathcal{V}_2$ other than B and Y (if $Y \in \mathcal{U}_2 \cup \mathcal{V}_2$). That is, Y is the first node in M_2 after B on a path from B to Y in M_1.

Definition 2 *[Strong Domination of Variables] Let $M_1 = ((\mathcal{U}_1, \mathcal{V}_1, \mathcal{R}_1), \mathcal{F}_1)$ and $M_2 = ((\mathcal{U}_2, \mathcal{V}_2, \mathcal{R}_2), \mathcal{F}_2)$. Let $Par_M(C)$ denote the variables that are parents of C in (the causal graph corresponding to) M.* M_1 strongly dominates M_2 with respect to C, *denoted $M_1 \succeq_C M_2$, if the following conditions hold:*

MI1$_{M_1,M_2,C}$. *The parents of C in M_2 are the immediate M_2-ancestors of C in M_1.*

MI2$_{M_1,M_2,C}$. *Every path from an exogenous variable to C in M_1 goes through a variable in $Par_{M_2}(C)$.*

MI3$_{M_1,M_2,C}$. *Let $X = ((\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2)) - \{C\}$. Then for all settings $\vec{x}$ of the variables in $\vec{X}$, all values c of C, all contexts $\vec{u}_1$ for M_1, and all contexts $\vec{u}_2$ for M_2, we have*

$$(M_1, \vec{u}_1) \models [\vec{X} \leftarrow \vec{x}](C = c) \text{ iff } (M_2, \vec{u}_2) \models [\vec{X} \leftarrow \vec{x}](C = c).$$

If MI1$_{M_1,M_2,C}$ holds and, for example, B is a parent of C in M_2, then there may be a variable B' on the path from

B to C in M_1. Thus, M_1 has in a sense more detailed information than M_2 about the causal paths leading to C. $\text{MI1}_{M_1,M_2,C}$ is not by itself enough to say that M_1 and M_2 agree on the causal relations for C. This is guaranteed by $\text{MI2}_{M_1,M_2,C}$ and $\text{MI3}_{M_1,M_2,C}$. $\text{MI2}_{M_1,M_2,C}$ says that the variables in $Par_{M_2}(C)$ screen off C from the exogenous variables in M_1. (Clearly the variables in $Par_{M_2}(C)$ also screen off C from the exogenous variables in M_2.) It follows that if $(M_1, \vec{u}_1) \models [Par_{M_2}(C) \leftarrow \vec{x}](C = c)$ for some context $\vec{u}_1$, then $(M_1, \vec{u}) \models [Par_{M_2}(C) \leftarrow \vec{x}](C = c)$ for all contexts $\vec{u}$ in M_1, and similarly for M_2. In light of this observation, it follows that $\text{MI3}_{M_1,M_2,C}$ assures us that C satisfies the same causal relations in both models. We write $M_1 \not\succeq_C M_2$ if any of the conditions above does not hold.

Two technical observations: First, note that there is an abuse of notation in the statement of $\text{MI3}_{M_1,M_2,C}$. We allow the set $\vec{X}$ in the statement of $\text{MI3}_{M_1,M_2,C}$ to include exogenous variables. However, in giving the semantics of the causal language, we consider only formulas of the form $[\vec{X} \leftarrow \vec{x}]\varphi$ where $\vec{X}$ mentions only endogenous variables. (Note that it is possible that some variables that are exogenous in M_1 may be endogenous in M_2, and vice versa.) Suppose that $\vec{X} \cap \mathcal{U}_1 = \mathcal{U}_1'$ and $\vec{X}' = \vec{X} - \mathcal{U}_1'$; then by $(M_1, \vec{u}_1) \models [\vec{X} \leftarrow \vec{x}](C = c)$ we mean $(M_1, \vec{u}_1') \models [\vec{X}' \leftarrow \vec{x}'](C = c)$, where $\vec{x}'$ is $\vec{x}$ restricted to the variables in $\vec{X}'$, and $\vec{u}_1'$ agrees with $\vec{u}_1$ on the variables in $\mathcal{U}_1 - \mathcal{U}_1'$, and agrees with $\vec{x}$ on the variables in $\mathcal{U}_1'$. Second, despite the suggestive notation, $\succeq_C$ is not a partial order. In particular, it is not hard to construct examples showing that it is not transitive. However, $\succeq_C$ is a partial order on compatible models (see the proof of Proposition 4.1), which is the only context in which we are interested in transitivity, so the abuse of notation is somewhat justified.

Note that we have a relation $\succeq_C$ for each variable C that appears in both M_1 and M_2. Model M_1 may be more informative than M_2 with respect to C whereas M_2 may be more informative than M_1 with respect to another variable C'. Roughly speaking, M_1 and M_2 are *strongly compatible* if for each variable $C \in \mathcal{V}_1 \cap \mathcal{V}_2$, either $M_1 \succeq_C M_2$ or $M_2 \succeq_C M_1$. We then combine M_1 and M_2 by taking the equations for C to be determined by the model that has more information about C.

Example 1 *(Bradley, Dietrich, and List 2014) An aid agency consults two experts about causes of famine in a region. Both experts agree that the amount of rainfall (R) affects crop yield (Y). Specifically, a shortage of rainfall leads to poor crop yield. Expert 2 says that political conflict (P) can also directly affect famine. Expert 1, on the other hand, says that P affects F only via Y. The experts' causal graphs are depicted in Figure 1, where the graph on the left, M_1, describes expert 1's model, while the graph on the right, M_2, describes expert 2's model. These graphs already appear in the work of BDL. In these graphs (as in many other causal graphs in the literature), the exogenous variables are omitted; all the variables are taken to be endogenous. Neither $\text{MI1}_{M_1,M_2,F}$ nor $\text{MI1}_{M_2,M_1,F}$, since P is not an M_2-immediate ancestor of F in M_1. Similarly, neither $\text{MI1}_{M_1,M_2,Y}$ nor $\text{MI1}_{M_2,M_1,Y}$ holds, since P is not an M_1-immediate ancestor of Y in M_2 (indeed, it is not an ancestor at all). $\text{MI2}_{M_1,M_2,F}$ holds since every path in M_1 from an exogenous variable to F goes through a variable that is a parent of F in M_2 (namely, Y); $\text{MI2}_{M_2,M_1,F}$ does not hold (there is a path in M_2 to F via P that does not go through a parent of F in M_1). Although we are not given the equations, we also know that $\text{MI3}_{M_1,M_2,F}$ does not hold. Since P is a parent of F in M_2 according to expert 2, there must be a setting y of Y such that the value of F changes depending on the value of P if $Y = y$. This cannot be the case in M_1, since Y screens off P from F. It easily follows that taking $\vec{X} = (P, Y)$ we get a counterexample to $\text{MI3}_{M_1,M_2,F}$. Therefore, we have neither $M_1 \succeq_F M_2$ nor $M_2 \succeq_F M_1$.* ∎

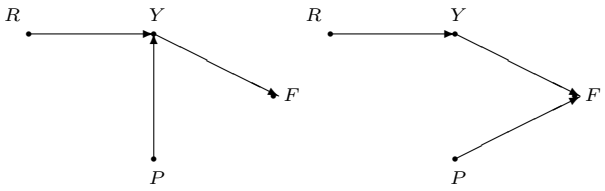

Figure 1: Two expert models of famine.

While the definition of dominance given above is useful, it does not cover all cases where we may want to combine models. Consider the following example, taken from the work of Sampson, Winship, and Knight (2013).

Example 2 *Two experts have provided causal models regarding the causes of domestic violence. According to the first expert, an appropriate arrest policy (AP) may affect both an offender's belief that his partner would report any abuse to police (PLS) and the amount of domestic violence (DV). The amount of domestic violence also affects the likelihood of a victim calling to report abuse (C), which in turn affects the likelihood of there being a random arrest (A). (Decisions on whether to arrest the offender in cases of domestic violence were randomized.)*

According to the second expert, DV affects A directly, while A affects the amount of repeated violence (RV) through both formal sanction (FS) and informal sanction on socially embedded individuals (IS). Sampson et al. (2013) use the following causal graphs shown in Figure 2, which are annotated with the direction of the influence (the only information provided by the experts) to describe the expert's opinions.

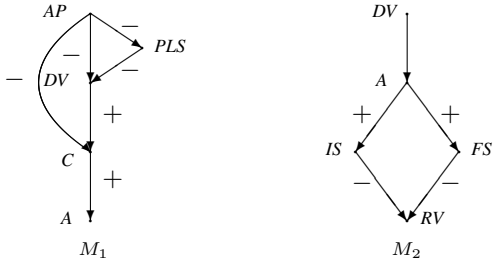

Figure 2: Expert's models of domestic violence.

For the two common variables DV and A, $\text{MI1}_{M_1,M_2,DV}$

and MI1$_{M_1,M_2,A}$ both hold. If the only variables that have exogenous parents are AP in M_1 and DV in M_2, and the set of parents of AP in M_1 is a subset of the set of parents of DV in M_2, then MI2$_{M_1,M_2,DV}$ holds. Sampson et al. seem to be implicitly assuming this, and that MI3 holds, so they combine M_1 and M_2 to get the causal graph shown in Figure 3.

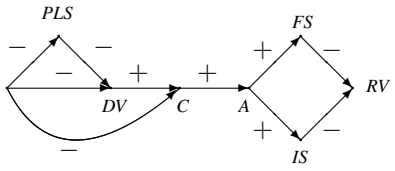

Figure 3: Combined experts' model of domestic violence.

Note that model M_2 in Figure 2 does not state how DV influences A. Presumably, this respresents the expert's understainty. We can capture this uncertainty by viewing the expert as having a probability on two models that disagree on the direction of DV's influence on A (and thus are incompatible because they disagree on the equations). We discuss in Section 5 how such uncertainty can be handled. ∎

Suppose that some parent of *AP* (or *AP* itself) in M_1 is not a parent of *DV* in M_2. Then, in M_1, it may be possible to change the value of *DV* by intervening on *AP*, while keeping the values of all the exogenous variables that are parents of *DV* in M_2 fixed. This will seem like an inexplicable change in the value of *DV* from the perspective of the second expert. If the second expert had been aware of such possible changes, she surely would have added additional variables to M_2 to capture this situation. One explanation of the fact that no changes were observed is that the second expert was working in a setting where the values of all variables that she cannot affect by an intervention are determined by some default setting of exogenous variables of which she is not aware (or not modeling). We now define a notion of domination that captures this intuition.

Definition 3 *[Weak Domination of Variable] Let $\vec{v}^*$ be a default setting for the variables in M_1 and M_2. M_1* weakly dominates *M_2 with respect to C relative to $\vec{v}^*$, denoted $M_1 \succeq_C^{\vec{v}^*} M_2$, if MI1$_{M_1,M_2,C}$ holds, and, in addition, the following condition (which can be viewed as a replacement for MI2$_{M_1,M_2,C}$ and MI3$_{M_1,M_2,C}$) holds:*

MI4$_{M_1,M_2,C,\vec{v}^*}$ *Let $\vec{X} = (\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2) - \{C\}$. Then for all settings $\vec{x}$ of the variables in $\vec{X}$, all values c of C, and all contexts $\vec{u}_1$ for M_1 and $\vec{u}_2$ for M_2 such that $\vec{u}_1$ and $\vec{u}_2$ agree on the variables in $\mathcal{U}_1 \cap \mathcal{U}_2$, $\vec{u}_1$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U}_1 - \mathcal{U}_2$, and $\vec{u}_2$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U}_2 - \mathcal{U}_1$, we have*

$$(M_1, \vec{u}_1) \models [\vec{X} \leftarrow \vec{x}](C = c) \text{ iff } (M_2, \vec{u}_2) \models [\vec{X} \leftarrow \vec{x}](C = c).$$

It is easy to see that $\succeq_C$ is a special case of $\succeq_C^{\vec{v}^*}$:

Lemma 1 *If $M_1 \succeq_C M_2$, then, for all default settings $\vec{v}^*$ of the variables in M_1 and M_2, we have $M_1 \succeq_C^{\vec{v}^*} M_2$.*

Proof: Suppose that $M_1 \succeq_C \mathcal{M}_2$. Fix default values $\vec{v}^*$. Clearly MI4$_{M_1,M_2,C,\vec{v}^*}$ is a special case of MI2. Thus, $M_1 \succeq_C^{\vec{v}^*} \mathcal{M}_2$. ∎

In light of Lemma 1, we give all the definitions in the remainder of the paper using $\succeq_C^{\vec{v}^*}$. All the technical results hold if we replace $\succeq_C^{\vec{v}^*}$ by $\succeq_C$ throughout.

Definition 4 *[Compatibility of Causal Models] If $M_1 = ((\mathcal{U}_1, \mathcal{V}_1, \mathcal{R}_1), \mathcal{F}_1)$ and $M_2 = ((\mathcal{U}_2, \mathcal{V}_2, \mathcal{R}_2), \mathcal{F}_2)$, then M_1 and M_2 are* compatible *if (1) for all variables $C \in (\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2)$, we have $\mathcal{R}_1(C) = \mathcal{R}_2(C)$ and (2) for all variables $C \in \mathcal{V}_1 \cap \mathcal{V}_2$, either $M_1 \succeq_C^{\vec{v}^*} M_2$ or $M_2 \succeq_C^{\vec{v}^*} M_1$. If M_1 and M_2 are compatible, then $M_1 \oplus M_2$ is the causal model $((\mathcal{U}, \mathcal{V}, \mathcal{R}), \mathcal{F})$, where*

- *$\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2 - (\mathcal{V}_1 \cup \mathcal{V}_2)$;*
- *$\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$;*
- *if $C \in \mathcal{U}_1 \cup \mathcal{V}_1$, then $\mathcal{R}(C) = \mathcal{R}_1(C)$, and if $C \in \mathcal{U}_2 \cup \mathcal{V}_2$, then $\mathcal{R}(C) = \mathcal{R}_2(C)$;*
- *if $C \in \mathcal{V}_1 - \mathcal{V}_2$ or if both $C \in \mathcal{V}_1 \cap \mathcal{V}_2$ and $M_1 \succeq_C^{\vec{v}^*} M_2$, then $\mathcal{F}(C) = \mathcal{F}_1(C)$; if $C \in \mathcal{V}_2 - \mathcal{V}_1$ or if both $C \in \mathcal{V}_1 \cap \mathcal{V}_2$ and $M_2 \succeq_C^{\vec{v}^*} M_1$, then $\mathcal{F}(C) = \mathcal{F}_2(C)$.*[2]

Returning to Example 2, assume that either MI2$_{M_1,M_2,DV}$, MI2$_{M_1,M_2,A}$, MI3$_{M_1,M_2,DV}$, and MI3$_{M_1,M_2,A}$ all hold, or there is a default setting $\vec{v}^*$ such that MI4$_{M_1,M_2,DV,\vec{v}^*}$ and MI4$_{M_1,M_2,A,\vec{v}^*}$ hold. Then $M_1 \oplus M_2$ has the causal graph described in Figure 3; that is, even though Sampson et al. (2013) do not have a formal theory for combining models, they actually combine models in just the way that we are suggesting.

Let $M_1 \sim_C^{\vec{v}^*} M_2$ be an abbreviation for $M_1 \succeq_C^{\vec{v}^*} M_2$ and $M_2 \succeq_C^{\vec{v}^*} M_1$. We also write $M_1 \succ_C^{\vec{v}^*} M_2$ if $M_1 \succeq_C^{\vec{v}^*} M_2$ and $M_2 \not\succeq_C^{\vec{v}^*} M_1$.

The next proposition provides evidence that Definition 4 is reasonable and captures our intuitions. Part (b) says that it is well defined, so that in the clauses in the definition where there might be potential conflict, such as in the definition of $\mathcal{F}(C)$ when $C \in \mathcal{V}_1 \cap \mathcal{V}_2$ and $M_1 \sim_C^{\vec{v}^*} M_2$, there is in fact no conflict; part (a) is a technical result needed to prove part (b). Part (c) states that the combined model in the compatible case is guaranteed to be acyclic. Part (d) says that causal paths in M_1 are preserved in $M_1 \oplus M_2$, while part (e) says that at least as far as formulas involving the variables in M_1 go, $M_1 \oplus M_2$ and M_1 agree, provided that the default values are used for the exogenous variables not in $\mathcal{U}_1 \cap \mathcal{U}_2$. Parts (d) and (e) can be viewed as saying that the essential causal structure of M_1 is preserved in $M_1 \oplus M_2$. (In the language of (**?**), part (e) says that $M_1 \oplus M_2$ is essentially a conservative extension of M_1.) Finally, parts (f) and (g) say that $\oplus$ is commutative and associative over its domain.

Proposition 4.1 *Suppose that M_1 M_2, and M_3 are pairwise compatible. Then the following conditions hold.*

[2] We are abusing notation here and viewing $\mathcal{F}_i(C)$ as a function from the values of the parents of C in M_i to the value of C, as opposed to a function from all the values of all variables other than C to the value of C.

(a) *If* $M_1 \sim_C^{\vec{v}^*} M_2$ *then (i)* $Par_{M_1}(C) = Par_{M_2}(C)$ *and (ii)* $\mathcal{F}_1(C) = \mathcal{F}_2(C)$*;*

(b) $M_1 \oplus M_2$ *is well defined.*

(c) $M_1 \oplus M_2$ *is acyclic.*

(d) If A *and* B *are variables in* M_1*, then* A *is an ancestor of* B *in* M_1 *iff* A *is an ancestor of* B *in* $M_1 \oplus M_2$*.*

(e) If φ *is a formula that mentions only the endogenous variables in* M_1*,* $\vec{u}$ *is a context for* $M_1 \oplus M_2$*,* $\vec{u}_1$ *is a context for* M_1*,* $\vec{u}$ *and* $\vec{u}_1$ *agree on the variables in* $\mathcal{U}_1 \cap \mathcal{U}_2$*,* $\vec{u}$ *agrees with* $\vec{v}^*$ *on the variables in* $\mathcal{U} - (\mathcal{U}_1 \cap \mathcal{U}_2)$*, and* $\vec{u}_1$ *agrees with* $\vec{v}^*$ *on the variables in* $\mathcal{U}_1 - \mathcal{U}_2$*, then* $(M_1, \vec{u}_1) \models \varphi$ *iff* $(M_1 \oplus M_2, \vec{u}) \models \varphi$*.*

(f) $M_1 \oplus M_2 = M_2 \oplus M_1$*.*

(g) If M_3 *is compatible with* $M_1 \oplus M_2$ *and* M_1 *is compatible with* $M_2 \oplus M_3$*, then* $M_1 \oplus (M_2 \oplus M_3) = (M_1 \oplus M_2) \oplus M_3$*.*

The proof is rather involved, and appears in full in Appendix 8.

We define what it means for a collection $\mathcal{M} = \{M_1, \ldots, M_n\}$ of causal models to be *mutually compatible* by induction on the cardinality of $\mathcal{M}$. If $|\mathcal{M}| = 1$, then mutual compatibility trivially holds. If $|\mathcal{M}| = 2$, then the models in $\mathcal{M}$ are mutually compatible if they are compatible according to Definition 4. If $|\mathcal{M}| = n$, then the models in $\mathcal{M}$ are mutually compatible if the models in every subset of $\mathcal{M}$ of cardinality $n-1$ are mutually compatible, and for each model $M \in \mathcal{M}$, M is compatible with $\oplus_{M' \neq M} M'$. By Proposition 4.1, if $M_1, \ldots, M_n$ are mutually compatible, then the causal model $M_1 \oplus \cdots \oplus M_n$ is well defined; we do not have to worry about parenthesization, nor the order in which the settings are combined. Thus, the model $\oplus_{M' \neq M} M'$ considered in the definition is also well defined. Proposition 4.1(e) also tells us that $M_1 \oplus \cdots \oplus M_n$ contains, in a precise sense, at least as much information as each model M_i individually. Thus, by combining mutually compatible models, we are maximizing our use of information.

We now discuss the extent to which our approach to combining models M_1 and M_2 satisfies BDL's desiderata. Recall that BDL considered only causal networks, not causal models in our sense; they also assume that all models mention the same set of variables. They consider four desiderata. We briefly describe them and their status in our setting.

- *Universal Domain*: the rule for combining models accepts all possible inputs and can output any logically possible model. This clearly holds for us.
- *Acyclicity*: the result of combining M_1 and M_2 is acyclic. This follows from Proposition 4.1(c), provided that $M_1 \oplus M_2$ is defined.
- *Unbiasedness*: if $M_1 \oplus M_2$ is defined, and M_1 and M_2 mention the same variables, then whether B is a parent of C in $M_1 \oplus M_2$ depends only on whether B is a parent of C in M_1 and in M_2, and This is trivial for us, since if B and C are in both M_1 and M_2 and $M_1 \oplus M_2$ is defined, then B is a parent of C in $M_1 \oplus M_2$ iff B is a parent of C in both M_1 and M_2. (The version of this requirement given by BDL does not say "if $M_1 \oplus M_2$ is defined", since they assume that arbitrary models can be combined.) BDL also have a *neutrality* requirement as part of unbiasedness. Unfortunately, an aggregation rule that says that B is a parent of C in $M_1 \oplus M_2$ iff B is a parent of C in both M_1 and M_2 (which seems quite reasonable to us) is not neutral in their sense, so we do not satisfy neutrality.
- *Non-Dictatorship:* no single expert determines the aggregation. This clearly holds for us.

4.2 Partial Compatibility

While the notion of dominance used in Definition 4 is useful, it still does not cover many cases of interest. Although Useem and Clayton do not provide causal models, we construct these based on the description provided. We do not provide a detailed explanation of all the variables and their dependencies here (details are provided in the full paper); for our purposes, it suffices to focus on the structure of these models.

Example 3 *Consider the two causal models in Figure 4. The SCICH model represents expert 1's opinion about emerging radicalization (R) in the State Correctional Institution Camp Hill in Pennsylvania. The TDCR model represents expert 2's opinion about the causes of emergence in the Texas Department of Corrections and Rehabilitation (TDCR). Both experts agree on the structural equations for R. However they differ on the structural equations for PD, CB and AM. The authors point to three main factors upon which the emergence of radicalization settings in both prison settings is dependent: "order in prisons" (PD), "a boundary between the prison and potentially radicalizing communities" (CB), and "having missionary leadership within the prison organizations" (AM). The also both share the same outcome —emerging radicalization (R). As can be observed from the descriptions provided, some variables and their dependency relations are specific to a particular prison. In the SCICH case, PD was attributed to corruption (CG) and lax management (LM) in the prison's staff together with prisons being allowed to roam freely (FM). CB was seen to be a result of religious leaders within the facilities being permitted to provide religious services freely (IL) and by prisoners showing membership within a prison community (CM) which in turn was signalled by prisoners allowed to wear distinguished street clothing (SC). Prison authorities' exercising of internal punishments, such as administrative segregation (AS), away from external oversight, and IL were considered to directly contribute to AM. TDCR instead considered PD to be linked to the rapid growth in inmates numbers (RG), inmates being allowed to assist prison authorities in maintaining order (T) and inmates feeling significantly deprived (D) within the prisons—the latter as a result of being forced to engage in unpaid work (W) and having limited contact with visitors (C). For the common variables CB and AM, assuming default values for CG, LM and FM in SCICH and for T, D and RG in TDCR, we can show that SCICH* $\succeq_{CB}^{\vec{v}^*}$ *TDCR and SCICH* $\succeq_{AM}^{\vec{v}^*}$ *TDCR. However, neither model dominates the other with respect to PD; neither* $MI1_{SCICH,TDCR,PD}$ *nor* $MI1_{TDCR,SCICH,PD}$ *holds. Therefore the models are incompatible according to Definition 4.*

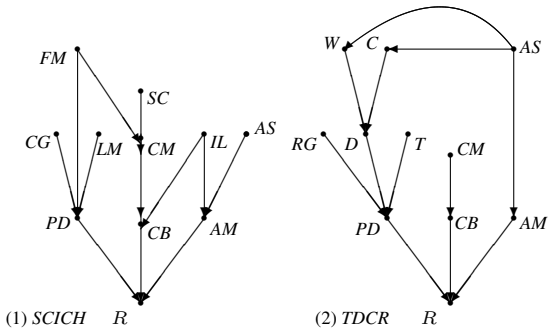

Figure 4: Schematic representation of the two prison models.

Although the models are incompatible according to our definition, the incompatibility is "localized" to the variable *PD*. Moreover, it is not even clear that there is disagreement with regard to *PD*; the experts could just be focusing on different variables. In a richer model, *PD* might have six parents. The trouble is, we have no idea from the two models what the equations for *PD* would be in the richer model.

Say that M_1 *weakly dominates* M_2 *with respect to a node* C *and default setting* v^*, written $M_1 \succeq_C^{v^*} M_2$ if $\text{MI4}_{M_1,M_2,C,\vec{v}^*}$ and the following weakening of $\text{MI1}_{M_1,M_2,C}$ hold:

MI1′$_{M_1,M_2,C}$ If A is a node in both M_1 and M_2 then A is an immediate M_1-ancestor of C in M_2 iff A is a parent of C in M_1.

Note that in Example 1, neither M_1 nor M_2 weakly dominates the other with respect to F: P is a parent of F in M_2 and is not an immediate M_1 ancestor of F in M_1, so M_1 does not weakly dominate M_2 with respect to F, while P is an immediate M_1 ancestor of F in M_2 and is not a parent of F in M_1, so M_2 does not weakly dominate M_1 either. Also note that MI1 implies MI1′; MI1′ is a strictly weaker condition than MI1, since it allows M_1 to weakly dominate M_2 with respect to C if C has parents in M_1 that are not in M_2 at all.

M_1 and M_2 are *weakly compatible* iff for all nodes in C in both $M_{\succeq_1}$ and M_2, either $M_1 \succeq_C^{v^*} M_2$ or $M_2 \succeq_C^{v^*} M_1$. Note that the two models in Figure 4 are weakly compatible.

If M_1 and M_2 are weakly compatible, then we can define $M_1 \oplus M_2 = ((\mathcal{U}, \mathcal{V}, \mathcal{R}), \mathcal{F})$ as follows:

- $\mathcal{U}, \mathcal{V}, \mathcal{R}$ are defined just as in Definition 4.
- For $\mathcal{F}$, if $C \in V_1 - V_2$ then $\mathcal{F}(C) = \mathcal{F}_1(C)$. and if $C \in V_2 - V_1$ then $\mathcal{F}(C) = \mathcal{F}_2(C)$. If $C \in V_1 \cap V_2$ and $M_1 \succeq_C^{v^*} M_2$, then let $\vec{P}_1$ consist of the parents of C in M_1 and let $\vec{P}_2$ consist of the parents of C in M_2 that are not in M_1. Then the parents of C in $M_1 \oplus M_2$ are the nodes $\vec{P}_1 \cup \vec{P}_2$. Let $\vec{v}_2$ consist be the values of the variables in $\vec{P}_2$ when the exogenous variables in M_2 have their default value in $\vec{v}^*$. Given an arbitrary setting $\vec{x}$ of the variables in $\vec{P}_1$, we define $\mathcal{F}(C)(\vec{x}, \vec{v}_2) = \mathcal{F}_2(C)(\vec{x})$. Symmetrically, if $C \in V_2 - V_1$ or both $C \in V_1 \cap V_2$ and $M_1 \succeq_C^{v^*} M_2$, then let $\vec{P}_1$ consist of the parents of C in M_1 and let $\vec{P}_2$ consist of the parents of C in M_2 that are not in M_1. Then the parents of C in $M_1 \oplus M_2$ are the nodes $\vec{P}_1 \cup \vec{P}_2$. Let $\vec{v}_2$ consist be the values of the variables in $\vec{P}_2$ when the exogenous variables in M_2 have their default value in $\vec{v}^*$. Given an arbitrary setting $\vec{x}$ of the variables in $\vec{P}_1$, we define $\mathcal{F}(C)(\vec{x}, \vec{v}_2) = \mathcal{F}_2(C)(\vec{x})$. We have a symmetric definition of $\mathcal{F}(C)$ if $C \in V_1 \cap V_2$ and $M_2 \succeq_C^{v^*} M_1$.

This definition does not define $\mathcal{F}(C)$ for all possible values of the parents of C. For example, if $C \in V_1 \cap V_2$ and $M_1 \succeq_C^{v^*} M_2$, we have not defined $\mathcal{F}(C)(\vec{x}, \vec{y})$ if $\vec{y}$ is a setting of the variables in $\vec{P}_2$ other than $\vec{v}_2$. Intuitively, this is because the experts have not given us the information to determine $\mathcal{F}(C)$ in these cases. We can think if $M_1 \oplus M_2$ as a *partial* causal model. Intuitively, we cannot define $\models$ in $M_1 \oplus M_2$ since we will not be able to define value of $(M_1 \oplus M_2, \vec{u}) \models C = c$ for some setting $\vec{u}$. Say that causal model $M^* = ((\mathcal{U}^*, \mathcal{V}^*, \mathcal{R}^*), \mathcal{F}^*)$ *extends* $M_1 \oplus M_2$ if $(\mathcal{U}^*, \mathcal{V}^*, \mathcal{R}^*) = (\mathcal{U}, \mathcal{V}, \mathcal{R})$ and $\mathcal{F}^*(C) = \mathcal{F}(C)$ whenever $\mathcal{F}(C)$ is defined. We now define a 3-valued version of $\models$ in $M_1 \oplus M_2$ by taking $(M_1 \oplus M_2, \vec{u}) \models \varphi$ iff $(M^*, \vec{u}) \models \varphi$ for all (complete) causal models M^* extending $M_1 \oplus M_2$ and taking $(M_1 \oplus M_2 \models \varphi$ to be *undefined* if neither $(M_1 \oplus M_2 \models \varphi$ nor $(M_1 \oplus M_2 \models \neg\varphi$ holds.

We can now prove a generalization of Proposition 4.1.

Proposition 4.2 *Suppose that M_1, M_2, and M_3 are pairwise weakly compatible. Then the following conditions hold.*

(a) If $M_1 \sim_C^{\vec{v}^*} M_2$ then (i) $Par_{M_1}(C) = Par_{M_2}(C)$ and (ii) $\mathcal{F}_1(C) = \mathcal{F}_2(C)$;

(b) $M_1 \oplus M_2$ is well defined.

(c) $M_1 \oplus M_2$ is acyclic.

(d) If φ is a formula that mentions only the endogenous variables in M_1, $\vec{u}$ is a context for $M_1 \oplus M_2$, $\vec{u}_1$ is a context for M_1, $\vec{u}$ and $\vec{u}_1$ agree on the variables in $\mathcal{U}_1 \cap \mathcal{U}_2$, $\vec{u}$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U} - (\mathcal{U}_1 \cap \mathcal{U}_2)$, and $\vec{u}_1$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U}_1 - \mathcal{U}_2$, then $(M_1, \vec{u}_1) \models \varphi$ iff $(M_1 \oplus M_2, \vec{u}) \models \varphi$.

(e) $M_1 \oplus M_2 = M_2 \oplus M_1$.

(f) If M_3 is weakly compatible with $M_1 \oplus M_2$ and M_1 is weakly compatible with $M_2 \oplus M_3$, then $M_1 \oplus (M_2 \oplus M_3) = (M_1 \oplus M_2) \oplus M_3$.

This approach to aggregating models is our main contribution. Using it, we show in the next section how experts' models can be combined to reason about interventions.

5 Combining Experts' Opinions

Suppose that we have a collection of pairs (M_1, p_1), ..., (M_n, p_n), with $p_i \in (0, 1]$; we can think of M_i as the model that expert i thinks is the right one and p_i as the

policymaker's degree of confidence that expert i is correct. Let $Compat = \{I \subseteq \{1, \ldots, n\}$: the models in $\{M_i : i \in I\}$ are mutually compatible$\}$. For $I \in Compat$, define $M_I = \oplus_{i \in I} M_i$. By Proposition 4.1, M_I is well defined. The policymaker considers the models in $\mathcal{M}_{Compat} = \{M_I : I \in Compat\}$, placing the probability of $p_I = \prod_{i \in I}(p_i) * \prod_{j \notin I}(1 - p_j)/N$ on M_I, where $N = \sum_{I \in Compat} p_I$ is a normalization factor.

Intuitively, we view the events "expert i is right" as being mutually independent, for $i = 1, \ldots, n$. Thus, p_I is the probability of the event that exactly the experts in I are right (and the ones not in I are wrong). If exactly the experts in I are indeed right, it seems reasonable to view M_I as the "right" causal model. Note that it is not possible for all the experts in I to be right if there are experts $i, j \in I$ such that M_i and M_j are incompatible. Thus, we consider only models M_I for $I \in Compat$. But even if $I \in Compat$, it is possible that some of the experts in I are wrong in their causal judgments. Our calculation implicitly conditions on the fact that at least one expert is right, but allows for the possibility that only some subset of the experts in I is right even if $I \in Compat$; we place positive probability on $M_{I'}$ even if I' is a strict subset of some $I \in Compat$. This method of combining experts' judgments is similar in spirit to the method proposed by Dawid (1987) and Fenton et al. (2016).

This completes our description of how to combine experts' causal judgments. At a high level, for each subset of experts whose judgments are compatible (in that the models they are proposing are pairwise compatible), we combine the models, and assign the combined model a probability corresponding the probability of the experts in the subset. Of course, once we have a probability on the settings in $\mathcal{M}_{Compat}$, we can compute, for each setting, which interventions affect the outcome φ of interest, and then compute the probability that a particular intervention is effective.

The straightforward strategy for a policymaker to compute the most effective intervention based on the experts' opinions and the degree of confidence of the policymaker in each expert's judgment is to compute the set $\mathcal{M}_{Compat}$ of models and then to apply the computation of interventions as described in Section 3 to each $M_I \in \mathcal{M}_{Compat}$. The probability that an intervention is effective is computed by summing the probability of the models where it is effective.

To get a sense of how this works, consider a variant of Example 1, in which a third expert provides her view on causes on famine and thinks that government corruption is an indirect cause via its effect on political conflict (see Figure 5); call this model M_3. According to the compatibility definition in Section 4, the models M_2 and M_3 are compatible (assuming that MI3 holds), but M_1 and M_3 are not. We have $\mathcal{M}_{Compat} = \{\{M_1\}, \{M_2\}, \{M_3\}, \{M_{2,3}\}\}$ with $M_{2,3} = M_2 \oplus M_3 = M_3$. Suppose that experts are assigned the confidence values as follows: $(M_1, 0.4)$, $(M_2, 0.6)$ and $(M_3, 0.5)$ respectively. Then the probability on $M_{2,3}$ is the probability of M_2 and M_3 being right (i.e., $0.6 * 0.5$) and M_1 being wrong (i.e., $1 - 0.4 = 0.6$). So we have $p_{2,3} = (0.6 * 0.5 * 0.6)/0.56 = 0.32$ (where 0.56 is the normalization factor). The probabilities on the other models is as follows.

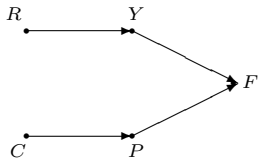

Figure 5: Third expert's (and combined) model of famine.

$$\begin{aligned} p_1 &= 0.4 * 0.4 * 0.5/0.56 = 0.14 \\ p_2 &= 0.6 * 0.6 * 0.5/0.56 = 0.32 \\ p_3 &= 0.6 * 0.4 * 0.5/0.56 = 0.21 \end{aligned}$$

The normalization factor N is simply $0.08 + 0.18 + 0.12 + 0.18 = 0.56$. Note that the number of models in $\mathcal{M}_{Compat}$ may be exponential in the number of experts. For example, if all experts are compatible, $Compat$ consists of all subsets of $\{1, \ldots, n\}$. The straightforward computation of interventions per model is exponential in the number of variables in the model. Since the number of variables in a combined model is at most the sum of the variables in each one, the problem is exponential in the number of experts and the total number of variables in the experts' models. In practice, however, we do not expect this to pose a problem. For the problems we are interested in, there are typically few experts involved; moreover, as we argued in Section 3, policymakers, in practice, restrict their attention to interventions on a small set of variables. Thus, we expect that the computation involved to be manageable.

Up to now, we have assumed that each expert proposes only one deterministic causal model. An expert uncertain about the model can propose several (typically incompatible) models, with a probability distribution on them. We can easily extend our framework to handle this. Suppose that expert i, with probability p_i of being correct, proposes m models $M_{i1}, \ldots, M_{im}$, where model M_{ij} has probability q_j of being the right one, according to i. To handle this, we simply replace expert i by m experts, $i_1, \ldots, i_m$, where expert i_j proposes model M_{ij} with probability $p_i q_j$ of being correct. As long as each of a few experts has a probability on only a few models, this will continue to be tractable.

6 Case Studies

In this section, we discuss the application of the framework to several case studies, demonstrating the concepts of compatibility and combinability and their effect on determining the best interventions.

6.1 Countering Domestic Child Abuse

We briefly discuss the relevant aspects of the case known as the "Baby P" case, which is a good illustration to the concepts of compatibility of experts' opinions regarding effective interventions. We then compare the Baby P case with another case of child abuse that resulted in child's death: Victoria Climbiè. As this case resembles the case of Baby P in many aspects, while being different in others, it is a good illustration of partial compatibility of models.

Baby P. "Baby P" (Peter Connelly) died in 2007 after suffering physical abuse over an extended period of time (**?**).

The court ultimately found the three adults living in a home with baby Peter guilty of "causing or allowing [Peter's] death" (**?**). After baby Peter's death, there was an extensive inquiry into practices, training, and governance in each of the involved professionals and organizations separately. In particular, an autopsy performed after Peter's death revealed that Peter's back had been broken, and he had sustained several other severe non-accidental injuries that were not detected by the doctor, because the doctor failed to perform a full examination. In the inquiry, England's health and social care regulator, the Care Quality Commission, criticized the hospital where the doctor worked for "poor recruitment practices", "lack of specific training in child protection", "shortages of staff", and "failings in governance" (**?**), all of which adversely affected the ability of the consultant paediatrician on duty to make sound decisions in this case. Since there there was no formal model of the case, some suggested interventions were later rejected. In particular, the head of children services at Haringey Council was removed from her post; she later won an appeal for unfair dismissal when the Court of Appeal deemed that the decision to dismiss her was not justified.

As shown by Chockler et al. 2015, the complete causal model for the Baby P case is complex, involving many variables and interactions between them. It is, however, decomposable into compatible submodels in the sense of Section 4. Specifically, we identify the submodels of "family life", "social services and police", "medical care", and "court". The schematic breakdown is presented in Figure 6.

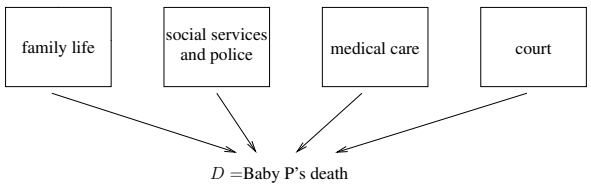

Figure 6: Schematic representation of causal sub-models in the baby P case.

There is exactly one expert for each submodel, each having degree of confidence 1. (Essentially, the investigation assumed that all experts were right.) The submodels overlap on some variables. First of all, they all have the variable D for death (whether or not baby P dies) and all the variables that are parents of D. Moreover, they agree on the equation that characterizes D in terms of its parents. We can partition the parents of D according to the submodel to which they are related. For example, all the parents of D not related to Baby P's family life are exogenous in the model corresponding to family life; all the parents of D not related to the court are exogenous in the model; and so on. More generally, if a variable is shared between two models, then it is exogenous in one and endogenous in the other, and no shared variable is an ancestor of another shared variable. Thus, MI1 and MI2 trivially hold for each shared variable (where the dominating model is the one where the variable is endogenous). Moreover, it is reasonable to assume that there is a default value such that MI4 holds. For example, the variable *CP* describes whether a child is put on the Child Protection Register. *CP* is endogenous in the social services and police submodel; its value is determined by criteria involving physical abuse and neglect. *CP* is exogenous is the medical care submodel. Its default value is 0 (the child is not on the Child Protection register). As long as the exogenous variables in social services and police submodel take on their default values (which involve children being well taken care of), *CP* will also take on value 0 in that submodel. Thus, MI4 holds for *CP*. Similar reasoning shows that MI4 holds for all variables. Thus, we can combine the submodels into one large model, which is effectively what Chockler et al. 2015 did in their analysis.

Victoria Climbiè. Victoria Climbiè's death occurred in 2000, seven years before Baby P, under the jurisdiction of the same council. Following her death from repeated abuse, an inquiry into her case suggested several interventions into the procedures of social workers and paediatricians. Yet, these interventions turned out to be inadequate, as the death of Baby P occurred under somewhat similar circumstances and his abuse also went unnoticed until his death.

Victoria Climbiè died in 2000, 18 months after arriving in the UK from the Ivory Coast to live with her great-aunt. Her great-aunt and the great-aunt's boyfriend were found guilty of Victoria's murder (in contrast with Baby P's case, where the adults in the house were found guilty of causing or allowing his death). The inquiry into the circumstances of Victoria's death placed the blame on social workers, who failed to notice Victoria's injuries, paediatricians, who accepted the explanation of Victoria's great-aunt that Victoria's injuries were self-inflicted, and the metropolitan police. In addition, the inquiry noted that the pastors in the church to which Victoria's great-aunt belonged, had concerns about the child's well-being but failed to contact any child protection services.

Victoria Climbiè died in her house from hypothermia, after sustaining a series of injuries over the 18 months that she was in custody of her great-aunt. As the circumstances of her death were different from that of Baby P (he died in the hospital), the causal model differs in the set of dependencies it captures. Victoria Climbiè's case is also complex with many variables and interactions between them, and, similarly to the Baby P's case, decomposable into compatible submodels in the sense of Section 4. The submodels are, however, a bit different. Specifically, we identify the submodels of "family life", "social services and police", "medical care", and "church" (note that this case was not presented in court prior to Victoria's death, so the submodel "court" is absent from the overall model).

The schematic breakdown is presented in Figure 7.

As evident from those and numerous other examples, case reviews and inquiries do not guarantee that similar cases will not happen in the future; in fact, Baby P died under the same council jurisdiction as Victoria Climbiè seven years earlier, despite interventions being proposed and implemented following the inquiry of the case. Following another case of a child dying from abuse in the hands of his family in 2013, Jones raised the question of whether the interventions implemented as a result of an inquiry are ever successful in

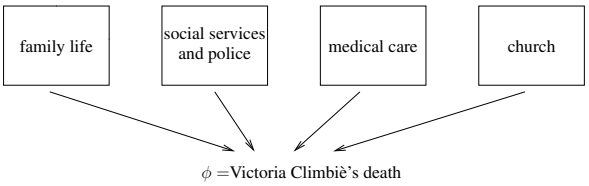

Figure 7: Schematic representation of causal sub-models in the Victoria Climbiè's case.

preventing similar cases in the future (Jones 2013). On the other hand, a child protection group stated, after reviewing both cases, that the interventions recommended in the case review of Victoria Climbiè's case had not been implemented, and that had they been implemented, it would prevent the death of Baby P in similar circumstances.

6.2 Countering Violent Behaviour in Prisons

Much of the recent literature on understanding and countering violent extremism (such as radicalization) has advocated for more structured data and analysis to support the making of policy interventions more effective (see, e.g., (Veldhuis and Kessels 2013)). We illustrate how our framework can help support this goal. Our case study elaborates on reports of inmate violence and radicalization in two US prison settings discussed by Useem and Clayton 2009, and the interventions that were introduced to tackle them. We note that radicalization in this context means, according to the authors, "[espousing] an ideology that endorses the use of violence calculated to spread fear, disrupt the social order, and achieve political goals external to the prison environment." The two models are presented graphically and briefly explained in Figure 4 of Example 3. Below, we explain in more detail the models and the meanings of the variables.

The first prison setting was that of the State Correctional Institution Camp Hill, near Graterford, Pennsylvania, denoted *SCICH*. In 1989, the facility witnessed a major riot resulting in damage to the buildings and injuries to 100 individuals. An investigation into the event showed evidence of corruption (*CG*) and lax management (*LM*) in the prison's staff causing disorder (*PD*) amongst inmates in the run-up to the riots. It was noted that, prior to the riot, prisoners were allowed to wear street clothing (*SC*) which, according to experts, the prisoners used to signal their membership in various communities within the prison (*CM*). In addition, prisoners were allowed to "move through the prison relatively freely" (*FM*), and prison authorities exercised internal punishments, such as administrative segregation (*AS*), away from external oversight. The second prison setting considered by Useem and Clayton was a Texas prison, denoted *TDCR*. It was noted that "rapid growth" (*RG*) in inmate numbers led to prison disorder (*PD*). Other influences on radicalization included inmates (1) being allowed to assist prison authorities in maintaining order, called the "building tenders" system (*T*) and (2) experiencing significantly greater deprivation *D*. The latter involved prisoners being required to work without being paid, being forced to wear short hair, and being given minimum allowance for phone calls. As in Graterford, the prison officials exercised administrative segregation, which contributed to the prisoners' sense of deprivation. Racial segregation (*RS*) contributed to increased membership in particular communities. The building tender system and racial segregation were the target of a court intervention. The court banned inmates from being involved in the enforcement of order within the prisons and ordered an end to racial segregation. These intervention were deemed effective in ensuring less radicalization within Texas prisons.

Two causal models were constructed, one for *SCICH* and another for *TDCR* from the description provided by Useem and Clayton. Although the two models use quite different variables, they both point to three main factors upon which the emergence of radicalization settings in both prison settings is dependent: "order in prisons" (*PD*), "a boundary between the prison and potentially radicalizing communities" (*CB*), and having "agency leadership ... infuse their agencies with an antiradicalization mission" (*AM*).[3] As can be observed from the descriptions provided, some variables and their dependency relations are specific to a prison. However, both descriptions involve the same outcome—emerging radicalization (R)—and assume that R has the same parents: *PD*, *CB*, and *AM*. Indeed, both models agree that the equation for R is $R = PD \wedge CB \wedge AM$. Furthermore, both models agree that *AS* affects prison order—in the first it affects *PD* directly whereas in the second it does so through prisoners' increased sense of deprivation *D*.

Note that there is still a sense in which the models are compatible. Intuitively, we can think of model M_2 as assuming that *T*, D and *RG* have some default value, while M_1 assumes that *CG*, *LM* and *FM* have some default value. In a richer model, *PD* might have six parents. This suggests the need for a yet more general notion of combination, which is defined in Section 4.2.

7 Conclusions

We have provided a method for combining causal models whenever possible, and used that as a basis for combining experts' causal judgments in a way that gets around the impossibility result of Bradley, Dietrich, and List (2014). We provided a gradual weakening of our definition of compatibility, allowing us to combine models that only agree on some of their parts. Our approach can be viewed as a formalization of what was done informally in earlier work (Chockler et al. 2015; Sampson, Winship, and Knight 2013). Our analysis of the case studies suggests that our approach can be applied in practice. We believe that using causal models as a way of formalizing experts' judgments, and then providing a technique for combining these judgments, will prove to be a powerful tool with which to approach the problem of finding the best intervention(s) that can be performed to ameliorate a situation.

Acknowledgments: We thank Noemie Bouhana, Frederick Eberhardt, and anonymous reviewers for useful comments. Joe Halpern's work was supported by NSF grants

[3] We treat the the fourth factor relating to the education profile as exogenous.

IIS-1703846 and IIS-1718108, AFOSR grant FA9550-12-1-0040, ARO grant W911NF-17-1-0592, and the Open Philanthropy project. Dalal Alrajeh's work was supported by MRI grant FA9550-16-1-0516.

References

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

Chockler, H.; Fenton, N. E.; Keppens, J.; and Lagnado, D. A. 2015. Causal analysis for attributing responsibility in legal cases. In *Proc. 15th International Conference on Artificial Intelligence and Law (ICAIL '15)*, 33–42.

Claassen, T., and Heskes, T. 2010. Learning causal network structure from multiple (in)dependence models. In *Proc. 5th 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 Conf. on Uncertainty in Artificial Intelligence (UAI 2012)*, 207–217.

Dawid, A. 1987. The difficulty about conjunction. *Journal of the Royal Statistical Society, Series D* 36:9197.

Fenton, N.; Neil, M.; and Berger, D. 2016. Bayes and the law. *Annual Review of Statistics and Its Application* 3:51–77.

Genest, C., and Zidek, J. V. 1986. Combining probability distributions: a critique and an annotated bibliography. *Statistical Science* 1(1):114–148.

Glymour, C., and Wimberly, F. 2007. Actual causes and thought experiments. In Campbell, J.; O'Rourke, M.; and Silverstein, H., eds., *Causation and Explanation*. MIT Press. 43–67.

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

Halpern, J. Y. 2015. A modification of the Halpern-Pearl definition of causality. In *Proc. 24th International Joint Conf. on Artificial Intelligence (IJCAI 2015)*, 3022–3033.

Halpern, J. Y. 2016. *Actual Causality*. MIT Press.

Hitchcock, C. 2001. The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy* XCVIII(6):273–299.

Hitchcock, C. 2007. Prevention, preemption, and the principle of sufficient reason. *Philosophical Review* 116:495–532.

Hyttinen, A.; Eberhardt, F.; and Jarvisalo, M. 2014. Constraint-based causal discovery: conflict resolution with answer set programming. In *Proc. 30th Conference on Uncertainty in Artificial Intelligence (UAI 2014)*, 340–349.

Jones, T. September 17, 2013. Daniel Pelka: Do serious case reviews work? http://www.bbc.co.uk/news/uk-england-24107377. Last accessed December 21, 2017.

Korb, K. B.; Hope, L. R.; Nicholson, A. E.; and Axnick, K. 2004. Varieties of causal intervention. In *Proc. 8th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence (PRICAI-04)*. 322–331.

Lewis, D. 2000. Causation as influence. *Journal of Philosophy* XCVII(4):182–197.

Lu, T.-C., and Druzdzel, M. J. 2002. Causal models, value of intervention, and search for opportunities. *Advances in Bayesian Networks: Studies in Fuzziness and Soft Computing* 146(30):121–135.

Papadimitriou, C. H., and Yannakakis, M. 1982. The complexity of facets (and some facets of complexity). *Journal of Computer and System Sciences* 28(2):244–259.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition.

Sampson, R. J.; Winship, C.; and Knight, C. 2013. Translating causal claims: Principles and strategies for policy-relevant criminology. *Criminology, Causality, and Public Policy* 12(4):587–616.

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 *Proc. 14th International Conf. on Artificial Intelligence and Statistics (AISTATS 2011)*, pp. 3–15.

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.

Useem, B., and Clayton, O. 2009. Radicalization of U.S. prisoners. *Criminology & Public Policy* 561–592.

Veldhuis. T. M. and Kessels, E. 2-13. Thinking before Leaping: The Need for More and Structural Data Analysis in Detention and Rehabilitation of Extremist Offenders. *The Hague: International Centre for Counter-Terrorism*.

Wikström, P.-O., and Bouhana, N. 2017. Analysing terrorism and radicalization: A situational action theory. In LaFree, G., and Freilich, J., eds., *The Encyclopaedia of the Criminology of Terrorism*. John Wiley and Sons.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

8 Proof of Proposition 4.1

Proof: For part (a), suppose that $M_1 \sim_C^{\vec{v}^*} M_2$ and, by way of contradiction, that $Par_{M_1}(C) \neq Par_{M_2}(C)$. We can assume without loss of generality that there is some variable $Y \in Par_{M_1}(C) - Par_{M_2}(C)$. Let $\vec{Z} = Par_{M_1}(C) - \{Y\}$. Since Y is a parent of C in M_1, there must be some setting $\vec{z}$ of the variables in $\vec{Z}$ and values y and y' for Y such that $F_C^1(y, \vec{z}) \neq F_C^1(y', \vec{z})$ in M_1, where $F_C^1 = \mathcal{F}_1(C)$. Suppose that $F_C^1(y, \vec{z}) = c$ and $F_C^1(y', \vec{z}) = c'$. Let $\vec{X} = ((\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2))$. By MI1, $(Par_{M_1}(C) \cup Par_{M_2}(C)) \subseteq \vec{X}$. Let $\vec{x}$ be a setting of the variables in $\vec{X} - \{C\}$ such that $\vec{x}$ agrees with $\vec{z}$ for the variables in $\vec{Z}$ and $\vec{x}$ assigns y to Y. Let $\vec{x}'$ be identical to $\vec{x}$ except that it assigns y' to Y. Since the values of the variables in $Par_{M_1}(C)$ determine the value of C in M_1, for all contexts $\vec{u}_1$ for M_1, we have $(M_1, \vec{u}_1) \models [\vec{X} \leftarrow \vec{x}](C = c)$ and $(M_1, \vec{u}_1) \models [\vec{X} \leftarrow \vec{x}'](C = c')$. Since $\vec{x}$ and $\vec{x}'$ assign the same values to all the variables in $Par_2(C)$, we must have $(M_2, \vec{u}_2) \models [\vec{X} \leftarrow \vec{x}](C = c)$ iff

$(M_2, \vec{u}_2) \models [\vec{X} \leftarrow \vec{x}'](C = c)$ for all contexts $\vec{u}_2$ for M_2. Thus, we get a contradiction to MI4$_{M_1,M_2,C}$. It follows that $Par_{M_1}(C) = Par_{M_2}(C)$. The fact that $\mathcal{F}_1(C) = \mathcal{F}_2(C)$ also follows from MI4$_{M_1,M_2,C}$. For suppose that $\vec{z}$ is a setting of the variables in $Par_1(C) = Par_2(C)$ and $\vec{x}$ is a setting of the variables in $\vec{X}' = \vec{X} - \{C\}$ that agrees with $\vec{z}$ on the variables in $Par_1(C)$. Then, for all contexts $\vec{u}_1$ for M_1 and $\vec{u}_2$ for M_2 such that $\vec{u}_1$ and $\vec{u}_2$ agree on the variables in $\mathcal{U}_1 \cap \mathcal{U}_2$, we have $F^1_C(\vec{z}) = c$ iff $(M_1, \vec{u}_1) \models [\vec{X}' \leftarrow \vec{x}](C = c)$ iff $(M_2, \vec{u}_2) \models [\vec{X}' \leftarrow \vec{x}](C = c)$ (by MI4$_{M_1,M_2,C}$) iff $F^2_C(\vec{z}) = c$. Thus, $\mathcal{F}_1(C) = \mathcal{F}_2(C)$.

For part (b), note that $M_1 \oplus M_2$ is well defined unless (i) $\mathcal{R}_1(C) \neq \mathcal{R}_2(C)$ for some $C \in ((\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2))$ or (ii) $M_1 \sim^{\vec{v}^*}_C M_2$ but $\mathcal{F}_1(C) \neq \mathcal{F}_2(C)$ for some $C \in \mathcal{V}_1 \cap \mathcal{V}_2$. Since M_1 and M_2 are compatible, (i) cannot happen; by part (a), (ii) cannot happen.

For part (c), we first show part (d): if A and B are both nodes in M_1 (i.e., A and B are in $\mathcal{U}_1 \cup \mathcal{V}_1$), then (the node labeled) A is an ancestor of (the node labeled) B in (the causal graph corresponding to) M_1 iff A is an ancestor of B in $M_1 \oplus M_2$, and similarly for M_2.

Suppose that A is an ancestor of B in M_1, Then there is a finite path $A_0, \ldots, A_n$ in the causal graph for M_1, where $A_0 = A$ and $A_n = B$. We first show that if $A_0, \ldots, A_n$ is an arbitrary sequence of nodes in M_1 such that none of the intermediate nodes (i.e., $A_1, \ldots, A_{n-1}$) are in M_2, and either $A_0 = A_n$ or at most one of A_0 and A_n is in M_2, then $A_0, \ldots, A_n$ is a path in M_1 iff $A_0, \ldots, A_n$ is a path in $M_1 \oplus M_2$. We proceed by induction on n, the length of the path. Since all the nodes in M_1 are nodes in $M_1 \oplus M_2$, the result clearly holds if $n = 0$. Suppose that $n > 0$ and the result holds for $n - 1$; we prove it for n. We cannot have $A_n \in \mathcal{U}_1 - \mathcal{V}_2$, since then A_n has no parents in M_1 or $M_1 \oplus M_2$. If $A_n \in \mathcal{V}_1 - \mathcal{V}_2$ or $A_n \in \mathcal{V}_1 \cap \mathcal{V}_2$ and $M_1 \succeq^{\vec{v}^*}_{A_n} M_2$, then $\mathcal{F}_{1,2}(A_n) = \mathcal{F}_1(A_n)$, so the parents of A_n in M_1 are also the parents of A_n in $M_1 \oplus M_2$. In particular, A_{n-1} is a parent of A_n in $M_1 \oplus M_2$ iff A_{n-1} is a parent of A_n in $M_1 \oplus M_2$, and the result follows from the induction hypothesis. Finally, if $A_n \in (\mathcal{U} \cup \mathcal{V}_1) \cap \mathcal{V}_2$ and $M_2 \succeq^{\vec{v}^*}_{A_n} M_1$, then $\mathcal{F}_{1,2}(A_n) = \mathcal{F}_2(A_n)$, so A_{n-1} must be in M_2. But this contradicts our assumption, that no intermediate nodes are in M_2 and at most one of A_0 and A_n is in M_2. This completes the argument. Note that the same argument applies if we reverse the roles of M_1 and M_2.

Now suppose that there are $m > 0$ nodes in M_2 on the path from A to B in M_1, say $C_1, \ldots, C_m$, in that order. We show that (i) C_m is an ancestor of B in $M_1 \oplus M_2$, (ii) A is an ancestor of C_1 in $M_1 \oplus M_2$, and (iii) C_1 is an ancestor of C_m in $M_1 \oplus M_2$. Parts (i) and (ii) follow from the earlier argument, since there are no intermediate nodes in M_2 on the path from C_m to B or on the path from A to C_1. So it remains to prove part (iii). We proceed by induction on m. If $m = 1$, the result is trivially true, since C_1 is a node in $M_1 \oplus M_2$. So suppose that $m > 1$. Since M_1 and M_2 are compatible and C_2 is a node in both M_1 and M_2 for $j > 1$, we must have either $M_1 \succeq^{\vec{v}^*}_{C_2} M_2$ or $M_2 \succeq^{\vec{v}^*}_{C_2} M_1$. If $M_1 \succeq^{\vec{v}^*}_{C_2} M_2$ then the parents of C_2 in M_1 are the parents of C_2 in $M_1 \oplus M_2$. In particular, if D is the parent of C_2 on the path from C_1 to C_2 in M_1, then D is a parent of C_2 in $M_1 \oplus M_2$. Since none of the intermediate nodes on the path from C_1 to D in M_1 are in M_2 except for C_1, it follows by our earlier argument than the path from C_1 to D in M_1 is also a path from C_1 to D in $M_1 \oplus M_2$. Thus, C_1 is an ancestor of C_2 in $M_1 \oplus M_2$. If $M_2 \succeq^{\vec{v}^*}_{C_2} M_1$, then the parents of C_2 in M_1 must also be in M_2 (in fact, they must be M_1-immediate ancestors of C_2 in M_2). Since none of the intermediate nodes on the path from C_1 to C_2 is in M_2, it must be the case that the path from C_1 to C_2 has length 1, and C_1 is a parent of C_2 in M_1. By MI1$_{M_2,M_1,C_2}$, there is a path from C_1 to C_2 in M_2 none of whose intermediate nodes is in M_1. Then the same argument given for the case that $M_1 \succeq^{\vec{v}^*}_{C_2} M_2$ shows that this path in M_2 also exists in $M_1 \oplus M_2$. Thus, C_1 is an ancestor of C_2 in $M_1 \oplus M_2$ in this case as well. The fact that C_2 is ancestor of C_m in $M_1 \oplus M_2$ follows from the induction hypothesis. Thus, C_1 is an ancestor of C_m in $M_1 \oplus M_2$.

For the converse, suppose that A and B are nodes in M_1 and A is an ancestor of B in $M_1 \oplus M_2$. We want to show that A is an ancestor of B in M_1. The argument is similar to that above, but slightly simpler. Again, there is a finite path $A_0, \ldots, A_n$ in the causal graph for $M_1 \oplus M_2$, where $A_0 = A$ and $A_n = B$. If none of the intermediate nodes on the path are in M_2 and at most one of A_0 and A_n is in M_2, then our initial argument shows that this path also exists in M_1.

Now suppose that there are $m > 0$ nodes in M_2 on the path from A to B in $M_1 \oplus M_2$, say $C_1, \ldots, C_m$, in that order. Much like before, we show that (i) C_m is an ancestor of B in M_1, (ii) A is an ancestor of C_1 in M_1, and (iii) C_1 is an ancestor of C_m in M_1. And again, parts (i) and (ii) follow from the earlier argument, since there are no intermediate nodes in M_2 on the path from C_m to B or the path from A to C_1. For part (iii), we again proceed by induction on m. If $m = 1$, the result is trivially true. So suppose that $m > 1$. Since M_1 and M_2 are compatible and C_2 is a node in both M_1 and M_2 for $j > 1$, we must have either $M_1 \succeq^{\vec{v}^*}_{C_2} M_2$ or $M_2 \succeq^{\vec{v}^*}_{C_2} M_1$. If $M_1 \succeq^{\vec{v}^*}_{C_2} M_2$, then the parents of C_2 in M_1 are just the parents of C_2 in $M_1 \oplus M_2$, so if D is the parent of C_2 on the path from C_1 to C_2 in $M_1 \oplus M_2$, D is a parent of C_2 in M_1. Since the path from C_1 to D in $M_1 \oplus M_2$ has no intermediate nodes in M_2, we can apply earlier argument to show that there is a path from C_1 to D in M_1, and complete the proof as before. If $M_2 \succeq^{\vec{v}^*}_{C_2} M_1$, then all the parents of C_2 in $M_1 \oplus M_2$ must be in M_2, so the path has length 1 and C_1 is a parent of C_2 in $M_1 \oplus M_2$ and in M_2. Thus, C_1 is an immediate M_1-ancestor of C_2 in M_2. MI1$_{M_2,M_1,C_2}$ implies that C_1 must be a parent of C_2 in M_1. Again, we can complete the proof as before.

The acyclicity of $M_1 \oplus M_2$ is now almost immediate. For suppose that there is a cycle $A_0, \ldots, A_n$ in the causal graph for $M_1 \oplus M_2$, where $A_0 = A_n$ and $n > 0$. Either A_n and A_{n-1} are both in M_1 (if $\mathcal{F}_{1,2}(A_n) = \mathcal{F}_1(A_n)$) or they are both in M_2 (if $\mathcal{F}_{1,2}(A_n) = \mathcal{F}_2(A_n)$). Suppose that they are both in M_1. Then, since A_{n-1} is an ancestor of A_n in $M_1 \oplus M_2$ and A_n is an ancestor of A_{n-1} in $M_1 \oplus M_2$, by the preceding argument, A_{n-1} is an ancestor of A_n in

M_1 and A_n is an ancestor of A_{n-1} in M_1, contradicting the acyclicity of M_1. A similar argument applies if both A_{n-1} and A_n are in M_2.

For part (e), suppose that $\vec{u}$ and $\vec{u}_1$ agree on the variables in $\mathcal{U}_1 \cap \mathcal{U}_2$, $\vec{u}$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U} - (\mathcal{U}_1 \cap \mathcal{U}_2)$, and $\vec{u}_1$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U}_1 - \mathcal{U}_2$. It clearly suffices to show that $(M_1, \vec{u}_1) \models \varphi$ iff $(M_1 \oplus M_2, \vec{u}) \models \varphi$ if φ has the form $[\vec{X} \leftarrow \vec{x}](Y = y)$, where $(\vec{X} \cup \{Y\}) \subseteq \mathcal{V}_1$. To show this, it suffices to show that $((M_1)_{\vec{X}=\vec{x}}, \vec{u}_1) \models (Y = y)$ iff $((M_1 \oplus M_2)_{\vec{X}=\vec{x}}, \vec{u}) \models (Y = y)$. Define the *depth* of a variable Y in a causal graph to be the length of the longest path from an exogenous variable to Y in the graph. We prove, by induction on the depth of the variable Y in the causal graph of $M_1 \oplus M_2$, that for all contexts $\vec{u}_1$ in M_1, $\vec{u}_2$ in M_2, and $\vec{u}$ in $M_1 \oplus M_2$, (i) if $Y \in \mathcal{U}_1 \cup \mathcal{V}_1$, $\vec{X} \subseteq \mathcal{V}_1$, $\vec{u}$ and $\vec{u}_1$ agree on the variables in $\mathcal{U}_1 \cap \mathcal{U}_2$, $\vec{u}$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U} - (\mathcal{U}_1 \cap \mathcal{U}_2)$, and $\vec{u}_1$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U}_1 - \mathcal{U}_2$, then $((M_1)_{\vec{X}=\vec{x}}, \vec{u}_1) \models (Y = y)$ iff $((M_1 \oplus M_2)_{\vec{X}=\vec{x}}, \vec{u}) \models (Y = y)$, and (ii) if $Y \in \mathcal{U}_2 \cup \mathcal{V}_2$, $\vec{X} \subseteq \mathcal{V}_2$, $\vec{u}$ and $\vec{u}_2$ agree on the variables in $\mathcal{U}_1 \cap \mathcal{U}_2$, $\vec{u}$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U} - (\mathcal{U}_1 \cap \mathcal{U}_2)$, and $\vec{u}_2$ agrees with $\vec{v}^*$ on the variables in $\mathcal{U}_2 - \mathcal{U}_1$, then $((M_2)_{\vec{X}=\vec{x}}, \vec{u}_2) \models (Y = y)$ iff $((M_1 \oplus M_2)_{\vec{X}=\vec{x}}, \vec{u}) \models (Y = y)$. (Note that if $Y \in (\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2)$, then it must satisfy both (i) and (ii).)

If Y has depth 0, then Y is an exogenous variable, and the result is immediate. Suppose that Y has depth $d > 0$. If $Y \in \mathcal{V}_1 - (\mathcal{U}_2 \cup \mathcal{V}_2)$, then the parents of Y in $M_1 \oplus M_2$ are the same as the parents of Y in M_1; (i) is then immediate from the induction hypothesis and (ii) is vacuously true. Similarly, if $Y \in \mathcal{V}_2 - (\mathcal{U}_1 \cup \mathcal{V}_1)$, then (ii) is immediate from the induction hypothesis and (i) is vacuously true. If $Y \in (\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2)$ and $M_1 \succeq_Y^{\vec{v}^*} M_2$, then again, the parents of Y in $M_1 \oplus M_2$ are the same as the parents of Y in M_1, so (i) is immediate from the induction hypothesis. To show that (ii) holds, fix appropriate contexts $\vec{u}_2$ and $\vec{u}$. Now the parents of Y in M_2 are the immediate M_2-ancestors of Y in M_1. Let $\vec{Z} = Par_{M_2}(Y)$. It follows from the arguments for part (c) that for all $Z \in Par_{M_2}(Y)$, all the paths from Z to Y in M_1 also exist in $M_1 \oplus M_2$ and the parents of Y in M_2 are exactly the immediate M_2-ancestors of Y in $M_1 \oplus M_2$. That is, $\vec{Z}$ screens Y from all other variables in M_2 not only in M_2, but also in M_1 and $M_1 \oplus M_2$. Suppose that $((M_2)_{\vec{X} \leftarrow x}, \vec{u}_2) \models \vec{Z} = \vec{z}$. It follows from the induction hypothesis that $((M_1 \oplus M_2)_{\vec{X}=\vec{x}}, \vec{u}) \models \vec{Z} = \vec{z}$. Let $\vec{W} = ((\mathcal{U}_1 \cup \mathcal{V}_1) \cap (\mathcal{U}_2 \cup \mathcal{V}_2)) - \{Y\}$. Let $\vec{w}$ be a setting for $\vec{W}$ that agrees with $\vec{z}$ on the variables in $\vec{Z}$. Then we have the following chain of equivalences:

$$
\begin{array}{ll}
 & ((M_2)_{\vec{X}=\vec{x}}, \vec{u}) \models Y = y \\
\text{iff} & ((M_2)_{\vec{X}=\vec{x}}, \vec{u}_2) \models [\vec{Z} \leftarrow \vec{z}](Y = y) \\
\text{iff} & ((M_2)_{\vec{X}=\vec{x}}, \vec{u}_2) \models [\vec{W} \leftarrow \vec{w}](Y = y) \\
\text{iff} & (M_2, \vec{u}_2) \models [\vec{W} \leftarrow \vec{w}](Y = y) \\
\text{iff} & (M_1, \vec{u}_1) \models [\vec{W} \leftarrow \vec{w}](Y = y) \quad \text{[by MI4}_{M_1, M_2, Y}\text{]} \\
\text{iff} & (M_1, \vec{u}_1) \models [\vec{Z} \leftarrow \vec{z}](Y = y) \\
\text{iff} & ((M_1)_{\vec{Z}=\vec{z}}, \vec{u}_1) \models (Y = y) \\
\text{iff} & ((M_1 \oplus M_2)_{\vec{Z}=\vec{z}}, \vec{u}_1) \models (Y = y) \quad \text{[already shown]} \\
\text{iff} & ((M_1 \oplus M_2), \vec{u}_1) \models [\vec{Z} \leftarrow \vec{z}](Y = y) \\
\text{iff} & ((M_1 \oplus M_2)_{\vec{X}=\vec{x}}, \vec{u}_1) \models [\vec{Z} \leftarrow \vec{z}](Y = y) \\
\text{iff} & ((M_1 \oplus M_2)_{\vec{X}=\vec{x}}, \vec{u}_1) \models Y = y \\
 & \quad \text{[since } (M_1 \oplus M_2)_{\vec{X}=\vec{x}}, \vec{u}_1) \models \vec{Z} = \vec{z}\text{]}
\end{array}
$$

This completes the proof of (d).

Part (f) is immediate from the definitions.

For part (g), suppose that $M_1 = ((\mathcal{U}_1, \mathcal{V}_1, \mathcal{R}_1), \mathcal{F}_1)$, $M_2 = ((\mathcal{U}_2, \mathcal{V}_2, \mathcal{R}_2), \mathcal{F}_2)$, $M_3 = ((\mathcal{U}_3, \mathcal{V}_3, \mathcal{R}_3), \mathcal{F}_3)$, $M_1 \oplus M_2 = ((\mathcal{U}_{1,2}, \mathcal{V}_{1,2}, \mathcal{R}_{1,2}), \mathcal{F}_{1,2})$, $M_2 \oplus M_3 = ((\mathcal{U}_{2,3}, \mathcal{V}_{2,3}, \mathcal{R}_{2,3}), \mathcal{F}_{2,3})$, $M_1 \oplus (M_2 \oplus M_3) = ((\mathcal{U}_{1,2,3}, \mathcal{V}_{1,2,3}, \mathcal{R}_{1,2,3}), \mathcal{F}_{1,2,3})$, and $(M_1 \oplus M_2) \oplus M_3 = ((\mathcal{U}'_{1,2,3}, \mathcal{V}'_{1,2,3}, \mathcal{R}'_{1,2,3}), \mathcal{F}'_{1,2,3})$. We want to show that $M_1 \oplus (M_2 \oplus M_3) = (M_1 \oplus M_2) \oplus M_3$. It is almost immediate from the definitions that $\mathcal{U}_{1,2,3} = \mathcal{U}'_{1,2,3}$, $\mathcal{V}_{1,2,3} = \mathcal{V}'_{1,2,3}$, and $\mathcal{R}_{1,2,3} = \mathcal{R}'_{1,2,3}$. To show that $\mathcal{F}_{1,2,3} = \mathcal{F}'_{1,2,3}$, we show that for all variables $C \in \mathcal{V}_{1,2,3}$, $\mathcal{F}_{1,2,3}(C) = \mathcal{F}'_{1,2,3}(C)$. We proceed by cases. If $C = \mathcal{V}_1 - (\mathcal{V}_2 \cup \mathcal{V}_3)$, then $C \notin \mathcal{V}_{2,3}$, so it is easy to check that $\mathcal{F}_{1,2,3}(C) = \mathcal{F}'_{1,2,3} = \mathcal{F}_1(C)$. Similarly, if $C \in \mathcal{V}_2 - (\mathcal{V}_1 \cup \mathcal{V}_3)$, then $\mathcal{F}_{1,2,3}(C) = \mathcal{F}'_{1,2,3} = \mathcal{F}_2(C)$, and if $C \in \mathcal{V}_3 - (\mathcal{V}_1 \cup \mathcal{V}_2)$, then $\mathcal{F}_{1,2,3}(C) = \mathcal{F}'_{1,2,3} = \mathcal{F}_3(C)$.

If $C \in (\mathcal{V}_1 \cap \mathcal{V}_2) - \mathcal{V}_3$, since M_1 and M_2 are compatible, either $M_1 \succeq_C^{\vec{v}^*} M_2$ or $M_2 \succeq_C^{\vec{v}^*} M_1$ (or both). If $M_1 \succeq_C^{\vec{v}^*} M_2$, then $\mathcal{F}_{1,2}(C) = \mathcal{F}_1(C)$, so $\mathcal{F}_{1,2,3}(C) = \mathcal{F}_1(C)$. Since $C \notin \mathcal{V}_3$, we have $\mathcal{F}_{2,3}(C) = \mathcal{F}_2(C)$. If we also have $M_2 \succeq_C^{\vec{v}^*} M_1$, then by (a), $\mathcal{F}_1(C) = \mathcal{F}_2(C)$, and it is easy to see that $\mathcal{F}'_{1,2,3}(C) = \mathcal{F}_1(C)$. Now suppose that $M_2 \not\succeq_C^{\vec{v}^*} M_1$. M_1 is compatible with $M_2 \oplus M_3$, we must have either $M_1 \succeq_C^{\vec{v}^*} M_2 \oplus M_3$ or $M_2 \oplus M_3 \succeq_C^{\vec{v}^*} M_1$. It is easy to see that since $M_2 \not\succeq_C^{\vec{v}^*} M_1$, we cannot have $M_2 \oplus M_3 \succeq_C^{\vec{v}^*} M_1$, so we must have $M_1 \succeq_C^{\vec{v}^*} M_2 \oplus M_3$. It follows that $\mathcal{F}'_{1,2,3}(C) = \mathcal{F}_1(C)$. The argument is similar if $C \in (\mathcal{V}_1 \cap V_3) - \mathcal{V}_2$ or $C \in (V_2 \cap V_3) - \mathcal{V}_2$.

Finally, suppose that $C \in (\mathcal{V}_1 \cap \mathcal{V}_2 \cap \mathcal{V}_3)$. We first show that $\succeq_C^{\vec{v}^*}$ is transitive when restricted to M_1, M_2, and M_3. For suppose that $M_1 \succeq_C^{\vec{v}^*} M_2$ and $M_2 \succeq_C^{\vec{v}^*} M_3$. If $M_1 \sim_C^{\vec{v}^*} M_2$ or $M_2 \sim_C^{\vec{v}^*} M_3$, then it is easy to see that $M_1 \succeq_C^{\vec{v}^*} M_3$. So suppose that $M_1 \succ_C^{\vec{v}^*} M_2$ and $M_2 \succ_C^{\vec{v}^*} M_3$. Since M_1 and M_3 are compatible, we must have either $M_1 \succeq_C^{\vec{v}^*} M_3$ or $M_3 \succeq_C^{\vec{v}^*} M_1$. Suppose by way of contradiction that $M_3 \succ_C^{\vec{v}^*} M_1$. Let $\vec{X}_1 = Par_{M_1}(C)$, $\vec{X}_2 = Par_{M_2}(C)$, and $\vec{X}_3 = Par_{M_3}(C)$. We now construct an infinite sequence of variables $A_0, A_1, \ldots$ such that each variable in the sequence is either in $\vec{X}_2 - \vec{X}_1$, $\vec{X}_3 - \vec{X}_2$, or $\vec{X}_1 - \vec{X}_3$, and if variable

A_n is in $\vec{X}_i - \vec{X}_j$, then the next variable is in $\vec{X}_j$ and there is a path in M_j from A_n to A_{n+1}. We proceed by induction. Since $M_1 \succ_C^{\vec{v}^*} M_2$, by MI1$_{M_1,M_2,C}$ there must be at least one variable in $A_0 \in \vec{X}_2 - \vec{X}_1$ and a path from Z_1 to C in M_1 that does not go through any other variables in $\vec{X}_2$. Since $\vec{X}_1$ screens C from all ancestors in M_1, this path must go through a variable $A_1 \in \vec{X}_1 - \vec{X}_2$. If $A_1 \in \vec{X}_3$, then it is in $\vec{X}_3 - \vec{X}_2$; if $A_1 \notin \vec{X}_3$, it is in $\vec{X}_1 - \vec{X}_3$. Either way, A_1 is an appropriate successor of A_0 in the sequence. The inductive step of the argument is identical; if $A_n \in \vec{X}_i - \vec{X}_j$, we use the fact that $M_j \succ_C^{\vec{v}^*} M_i$ to construct A_{n+1}. Note that, for all $n \geq 0$, since $A_n \in \vec{X}_i - \vec{X}_j$ and $A_{n+1} \in \vec{X}_j$, we must have $A_n \neq A_{n+1}$. Moreover, by the argument in the proof of (c) since there is a path from A_n to A_{n+1} in M_j, there must also be such a path in $M_1 \oplus (M_2 \oplus M_3)$. Since there are only finitely many variables altogether, there must be some N_1 and N_2 such that $A_{N_1} = A_{N_2}$. That means we have a cycle in $M_1 \oplus (M_2 \oplus M_3)$, contradicting (c).

Since $\succeq_C^{\vec{v}^*}$ is transitive and complete on $\{M_1, M_2, M_3\}$ (completeness says that for each pair, one of the two must be dominant), one of M_1, M_2, and M_3 must dominate the other two with respect to $\succeq_C^{\vec{v}^*}$. Suppose it is M_1. It is easy to see that $M_1 \oplus M_2 \succeq_C^{\vec{v}^*} M_3$ and $M_1 \succeq_C^{\vec{v}^*} (M_2 \oplus M_3)$. It then easily follows that $\mathcal{F}_{1,2,3}(C) = \mathcal{F}'_{1,2,3}(C) = \mathcal{F}_1(C)$. A similar argument holds if M_2 or M_3 is the model that dominates with respect to $\succeq_C^{\vec{v}^*}$. ∎