Essential Structure for Causal Models

Jenn McDonald

Abstract This paper introduces and defends a new principle for when a structural equation model is *apt* for analyzing actual causation. Any such analysis in terms of these models has two components: a recipe for reading claims of actual causation off an *apt* model, and an articulation of what makes a model apt. The primary focus in the literature has been on the first component. But the recently discovered problem of structural isomorphs has made the second especially pressing (Hall 2007; Hitchcock 2007a). Those with realist sympathies have reason to resist the standard response to this problem, which introduces a normative parameter into the metaphysics (Gallow 2021; Hall 2007; Halpern 2016b; Halpern and Hitchcock 2010, 2015; Menzies 2017). However, the only alternative solution in the literature leaves central questions unanswered (Blanchard and Schaffer 2017). I propose an independently motivated aptness requirement, *Evident Mediation*, that provides the missing details and resolves the structural isomorph problem without need for a normative parameter.

§1 Introduction

A precise analysis of actual causation has proven elusive. But recent progress seems to have been made utilizing the framework of structural equation models. A *structural equation model* – or *causal model* – is a set of variables that represent the causal relata and a set of functional equations defined over them that represent dependency relations holding between the relata. There are two components working together in such an analysis. The first gives a recipe by which one can read actual causal relations off of an apt model. This answers the question of *which properties* of a model are the right ones for identifying a causal relation. Although this question has received considerable attention, it remains less than entirely settled. [1]

This paper focuses, though, on the second component. This answers the question of what certifies a particular interpreted causal model for use as one to which we can apply the correct recipe and get only true claims about what causes what in a target situation. What makes an interpreted causal model *apt*?[2] Answering this question has become all the more pressing due to the recently discovered *problem of structural isomorphs*.[3] One model can accurately represent two different situations, in the sense that the values of its variables are correct and all the counterfactuals it entails are true of both situations, and yet our judgment of what causes what differs in the two situations. Cases like these suggest that more than accuracy is needed to render an interpreted model *apt* for any given situation.

What we need, then, is some way to rule such models inapt for representing one (or both) of the situations. But there are reasons to resist the typical way of doing this – distinguishing between default and deviant states of a system (Gallow 2021; Hall 2007;

[1] See (Beckers and Vennekens 2017, 2018; Gallow 2021; Hall 2007; Halpern 2016a; Halpern and Pearl 2005; Hitchcock 2001b; Weslake 2015; Woodward 2003) for various causal model analyses of actual causation.

[2] For work on the question of when a model is apt, see (Blanchard and Schaffer 2017; Halpern 2016b; Halpern and Hitchcock 2010; Hitchcock 2001b, 2012; Woodward 2016).

[3] Also called the problem of *counterfactual* isomorphs. See (Blanchard and Schaffer 2017; Hall 2007; Hiddleston 2005b; Hitchcock 2007a; Menzies 2017).

Halpern 2016b; Halpern and Hitchcock 2010, 2015) – which I discuss in §2.3. Blanchard and Schaffer (2017) present the only alternative solution in the way of an aptness requirement, Essential Structure, which enjoins us to include "enough events to capture the essential structure of the situation being modelled (2017:183)" As is, however, this is inadequate in leaving unclear what structure counts as essential and thus leaving open whether the underlying rationale for this requirement can be given in objective terms.

This paper aims to rectify these inadequacies. I begin by uncovering and clarifying what underlies Essential Structure. I argue that it is the need to be sensitive to the presence of *partially mediating variables*, which I define. By requiring that partially mediating variables be explicitly included in a model, the problem of structural isomorphs dissolves. I call this new aptness principle *Evident Mediation* and argue that it is motivated independently of this particular problem. I thus propose that Evident Mediation be adopted as an improvement on and replacement of Essential Structure.

§2 The Problem of Structural Isomorphs

2.1 A SEM in Three Parts

A SEM has three parts: a signature, an assignment, and linkage.[4] I will describe these in turn, alongside how and what each is standardly taken to represent.

[4] The formalism of a SEM comes in many varieties. The formalism I present here is for the purposes of representing *particular* situations, rather than type-level ones, and follows the framework found in Halpern (2000) and Blanchard and Schaffer (2017).

First, the *signature* of a model is an ordered triple, $\boldsymbol{S} = <\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{R}>$, comprised of a set of *exogenous variables*, $\boldsymbol{U}$, which are the independent variables, a set of *endogenous variables*, $\boldsymbol{V}$, which are the dependent ones, and a relation, $\boldsymbol{R}$, that maps each variable onto a range of values with at least two members. Intuitively, variables represent causal relata. A particular factor in the target situation will be represented by a value of a variable, with the other values of that same variable representing alternatives. The term "factor" refers to any kind of thing that can be reasonably considered a relatum of causation.[5] The variables of a SEM can be used to represent any possible kind of relata – such as events, property instances, facts, etc. To simplify, I will use them only to represent events, and will adopt an exemplification theory of events. This takes events to be, roughly, an object exemplifying a property in a particular time period (Hendrickson 2006; Kim 1976), although I will leave the time period component implicit.[6] An example will illuminate. Consider:

> **Overdetermination** Suzy and Billy each throw a rock at a window, at the same time and with the same velocity. The rocks simultaneously arrive at the window, and the window shatters.

[5] This use of 'factor' follows Eells (1988, 1991), Hitchcock (2001a:362), and Menzies (2004b), among others.

[6] The proposal I make does not essentially rely on this view of events, but it strikes me as providing for the neatest exposition. For example, it allows us to treat all relata as of the same kind, even omissions (like Suzy *not* throwing). I take omissions to be events with a negative property as a component, bracketing controversy over the metaphysics of omissions as not directly bearing on this paper. See, among others, (Bernstein 2015; Henne, Pinillos, and De Brigard 2016; Schaffer 2004; Zangwill 2011).

We can represent this in a model, $\mathcal{M}_1$, with three binary variables, X, Y, and Z, each of which can take the value *0* or *1*, and each value is assigned an event description. In order to represent **Overdetermination**, we will interpret them so that $X = 1$ represents Suzy throwing her rock while $X = 0$ represents her not throwing, $Y = 1$ represents Billy throwing while $Y = 0$ represents him not throwing, and $Z = 1$ represents the window shattering while $Z = 0$ represents it not shattering.

$\mathcal{I}(\mathcal{M}_1)_O$: $$X = \begin{cases} 1 \; if \; Suzy \; throws \; a \; rock \\ 0 \; if \; Suzy \; doesn't \; throw \end{cases} \qquad Y = \begin{cases} 1 \; if \; Billy \; throws \; a \; rock \\ 0 \; if \; Billy \; doesn't \; throw \end{cases}$$

$$Z = \begin{cases} 1 \; if \; the \; window \; shatters \\ 0 \; if \; the \; window \; doesn't \; shatter \end{cases}$$

Here, the set of exogenous variables, $\boldsymbol{U}$, is the set containing only X and Y ($\boldsymbol{U} = \{X, Y\}$). This corresponds to our treating whether Suzy and Billy throws as independent variables in the situation. The set of endogenous variables, $\boldsymbol{V}$, is the set containing only Z ($\boldsymbol{V} = \{Z\}$). This corresponds to the fact that whether the window shatters is a function of whether either child throws. $\boldsymbol{R}$ maps each variable to the set of values, $\{0, 1\}$.

Second, the *assignment*, $\boldsymbol{\mathcal{A}}$, is a function that maps to each exogenous variable, $X \in \boldsymbol{U}$, one of its values, $x_1 \in \boldsymbol{R}(X)$. Intuitively, the assignment represents initial conditions. The assignment results in a set of fixed equations.[7] For example, the assignment that captures

[7] This is somewhat non-standard but formally benign, and in permitting interventions on exogenous variables allows for a cleaner presentation.

the initial conditions of **Overdetermination** on our interpretation is {$X = 1$, $Y = 1$}. This represents the event of Suzy and that of Billy actually throwing their rocks.

There is some freedom in selecting which events and which alternatives of those events to explicitly represent by a model. This is the process of *variable selection*. I will adopt the universal line of taking variable selection to be governed by at least three aptness principles: Exclusivity, Exhaustivity, and Distinctness. *Exclusivity* requires that a variable's values represent mutually exclusive events.[8] *Exhaustivity* requires that a variable's values exhaustively represent how the target event could have otherwise occurred.[9] Distinctness requires that different variables represent *distinct* events, understood along the lines of logical, conceptual, and mereological independence.[10] Call an interpretation *permissible* when what it says is exhaustive, exclusive, or distinct really *is* exhaustive, exclusive, or distinct in the target situation. Notice that in order for $\mathcal{I}(\mathcal{M}_1)_O$ to be considered *exhaustive* with respect to *X*, Suzy *could only have* thrown a rock. If it were possible that she throws a brick, for example, then X would not be exhaustive, rendering $\mathcal{I}(\mathcal{M}_1)_O$ impermissible. I will leave this kind of subtlety implicit by assuming whatever's required to make an interpretation permissible.

---

[8] See (Blanchard and Schaffer 2017:182; Briggs 2012:142; Hitchcock 2004:145, 2007b:76, 2007a:502; Pearl 2000:3; Woodward 2003:98).

[9] See (Blanchard and Schaffer 2017:182; Briggs 2012:142; Hitchcock 2001b:287; Pearl 2000:3; Woodward 2016:1064).

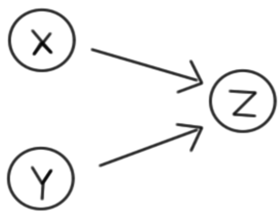
[10] See (Blanchard and Schaffer 2017; Briggs 2012; Hitchcock 2004:142, 2007a:502; Paul and Hall 2013:59; Woodward 2016:1063–64).

The final component of a model is the *linkage*, $\boldsymbol{\mathcal{L}}$, which is a set of functional equations defined over the variables from the signature. It stipulates the way in which the endogenous variables depend on the exogenous ones. Functional equations are asymmetric and minimal. They are *asymmetric* in that they stipulate what value the left-hand variable will take for any combination of values of the right-hand variables, when these variables are set to their values by *intervention*. An intervention on a variable is a surgical operation that forces *only* the specified variable to one of its specified values, and otherwise leaves the model as is. More precisely, an intervention, $I_{X=xi}$, on a variable, X, in a model produces a sub-model in which everything is the same as the original model except that the X-equation is replaced by '$X=x_i$'. [11] Such an operation renders X independent of its parent variables but preserves the dependency structure down stream of X. Equations are *minimal* in that the right-hand side of the equation includes only variables for which changes in their values result in a change in the left-hand variable when all other variables are held fixed.

$\boldsymbol{\mathcal{S}}$ = $\boldsymbol{U} = \{X, Y\}$
$\boldsymbol{V} = \{Z\}$
$\boldsymbol{R} = f(X_i) = \{1, 0\}$

$\boldsymbol{\mathcal{A}}$ = (EQ1) $X = 1$
(EQ2) $Y = 1$

$\boldsymbol{\mathcal{L}}$ = (EQ3) $Z := \max(X, Y)$



$\mathcal{M}_1$

The functional equations capture what actually happens in the case as well as what would have happened had the alternatives occurred instead. In **Overdetermination**, had either child thrown a rock the window would have shattered, and had neither child thrown it would not have shattered. This can be captured by the functional equation, $Z := max(X, Y)$.

[11] This follows (Pearl 2000). See also (Briggs 2012). For a different formalization see (Woodward 2003).

The literature divides on whether the dependencies represented by equations are counterfactual or type-level causal. Proponents of the former view generally continue the Lewisian tradition of seeking to reduce actual and type-level causal dependence to counterfactual dependence (Blanchard and Schaffer 2017; Hall 2007; Hitchcock 2001b). Proponents of the latter view treat type-level causal dependence (Halpern and Pearl 2005; Hiddleston 2005a; Pearl 2000; Woodward 2003) or causal/structural determination (Gallow 2016, 2021) as either primitive or at least prior, and generally seek to reduce both actual causal dependence and counterfactual dependence to the prior dependence. (The latter project being the development of an *interventionist semantics* for counterfactuals.) So, proponents of both views take a model under a given interpretation to *entail* a set of counterfactuals, just for different reasons. For proponents of the former view, counterfactuals are entailed because they are directly represented. For proponents of the latter, they're entailed because they supervene on the type-level causal structure that is represented.[12] As a result, I can speak of the counterfactuals *entailed* by a model under an interpretation and remain neutral on the underlying metaphysics. So, for example, the content of $\text{EQ3}_{\mathcal{M}_1}$ can be unpacked in the following entailed counterfactuals:

[12] I should flag that the causal dependence view will diverge from the counterfactual dependence view with respect to precisely which counterfactuals are entailed by a model-interpretation pair. This is because an interventionist semantics (aka. a causal model semantics) diverges from a similarity semantics in assignment of truth-values to counterfactuals (Woodward 2003), and in what counts as a valid inference (Briggs 2012). (Presumably, it also diverges from a premise semantics or filtering semantics, though there's no work on this to my knowledge.) Thus, the same model-interpretation pair will not entail the exact same counterfactuals given a causal interpretation as given a counterfactual one. I gloss over this complication since these divergences do not affect any of the counterfactuals used in this paper.

X = 1 (Suzy throws a rock) □→ *Z = 1* (window shatters).

Y = 1 (Billy throws a rock) □→ *Z =1* (window shatters).

X = 0 and *Y = 0* (Suzy nor Billy throws) □→ *Z = 0* (window doesn't shatter).

Given this, we can say what it is for a model on a permissible interpretation to be *accurate* of its intended situation: (i) for those values mapped to the exogenous variables by the assignment, their assigned event descriptions are realized by actual events; and (ii) the entailed counterfactuals are true. Notice that although $\mathcal{I}(\mathcal{M}_1)_O$ interprets *X = 1* as simply *Suzy throws a rock*, in order for EQ3 to entail true counterfactuals, Suzy's throw *must* be on the correct trajectory to hit the window and *must* be with sufficient force to shatter the window. Otherwise, it would be possible for Suzy to throw without the window shattering, rendering one of EQ3's entailed counterfactuals false. I will leave this kind of subtlety implicit by assuming whatever's required to make a model on its permissible interpretation accurate.

Two final points to make. The first is that for the purpose at hand, we focus on models that are *recursive* in the sense that the equations can be ordered such that once a variable appears on the right-hand side of an equation it does not again appear on the left-hand side. We assume a unique equation for each endogenous variable, and name each equation for its left-hand variable. So, call $\text{EQ3}_{\mathcal{M}_1}$ the *Z*-equation. The set of variables appearing on the right-hand side of the *Z*-equation are *Z*'s *parent* variables, with respect to which *Z* is a *child* variable.

Second and finally, it can be helpful to represent recursive structural equation models with directed acyclic graphs. These graphs abstract away from the precise nature of the represented dependencies, indicating only the fact that the dependencies exist. The variables of a SEM correspond to nodes and the equations correspond to directed edges between nodes, drawn as arrows from parent to child variables. (See $\mathcal{M}_1$ above.)

2.2 The Problem of Structural Isomorphs

Unfortunately, it often happens that one model can accurately represent two different situations – under two different interpretations, of course – and so delivers the same verdicts in both for what causes what. And yet our judgment of what causes what differs in the two situations. This is the *problem of structural isomorphs*. As an example, consider the following:

> **Bogus Prevention** Assassin intends to put poison in King's coffee, but at the very last-minute changes her mind and refrains. Bodyguard, though, compulsively puts antidote in King's coffee every morning, and had already done so this morning. King survives.[13]

Intuitively, Bodyguard's administration of antidote is not a cause of King surviving. King would have survived regardless of Bodyguard's compulsive action. Compare this with **Overdetermination**. While intuitions about symmetric overdetermination cases aren't as vivid, it is natural enough to take Suzy's throw and Billy's throw to be equal and

[13] This example originates with (Hiddleston 2005b:32). The exact version that I use comes from (Blanchard and Schaffer 2017:185; Hitchcock 2007a), and goes by the name "Careful Antidote" in (Weslake 2015:19). The general form is also found in Hall (2007:5.2), where it falls under the name "Non-Existent Threats."

independent actual causes of the window shattering. This means that Billy's throw is causally disanalogous to Bodyguard's administering antidote. However, **Bogus Prevention** can be represented accurately using *the very same model*, $\mathcal{M}_1$. We just need a different interpretation. Use $\mathcal{I}(\mathcal{M}_1)_{BP}$ to represent **Bogus Prevention**:

$\mathcal{I}(\mathcal{M}_1)_{BP}$: $$X := \begin{cases} 1 \; if \; Assassin \; doesn't \; administer \; poison \\ 0 \; if \; Assassin \; administers \; poison \end{cases}$$

$$Y := \begin{cases} 1 \; if \; Bodyguard \; administers \; antidote \\ 0 \; if \; Bodyguard \; doesn't \; administer \; antidote \end{cases} \qquad Z := \begin{cases} 1 \; if \; King \; survives \\ 0 \; if \; King \; dies \end{cases}$$

$\mathcal{M}_1$ is accurate of **Overdetermination** on $\mathcal{I}(\mathcal{M}_1)_O$ and of **Bogus Prevention** on $\mathcal{I}(\mathcal{M}_1)_{BP}$. The assignment accurately captures initial conditions of each, and the equations entail true counterfactuals. The assignment says truly of **Overdetermination** on $\mathcal{I}(\mathcal{M}_1)_O$ that Suzy throws and Billy throws, and it says truly of **Bogus Prevention** on $\mathcal{I}(\mathcal{M}_1)_{BP}$ that Assassin does not administer poison although Bodyguard does administer antidote. The counterfactuals entailed by $\mathcal{M}_1$ on $\mathcal{I}(\mathcal{M}_1)_O$ are true of **Overdetermination**, and those entailed by $\mathcal{M}_1$ on $\mathcal{I}(\mathcal{M}_1)_{BP}$ are true of **Bogus Prevention**. It is true of **Overdetermination** that had Suzy thrown, then the window would have shattered; had Billy thrown, then the window would have shattered; and had neither Suzy nor Billy thrown, then the window would not have shattered. Similarly, it is true of **Bogus Prevention** that had Assassin not administered poison, then King would have survived; had Bodyguard administered antidote, then King would have survived; and had Assassin administered poison but Bodyguard not administered antidote, then King would have died.

The crucial observation is that on the respective interpretations, Billy's throw is structurally analogous to Bodyguard's administration of antidote; *Y = 1* in each case. As a result, *any* recipe for actual causation applied to $\mathcal{M}_1$ that delivers the right results in one case will ipso facto deliver the wrong results in the other. The dilemma is that either our intuition is mistaken that these cases have a different causal structure, or else $\mathcal{M}_1$ is not suited on the respective interpretation for representing one or both cases. It is inapt, despite its accuracy.

2.3 Defaults and Deviants vs. Essential Structure

The literature has uniformly defended the second horn of this dilemma. The divisive question is: what else beyond accuracy does aptness require? The leading response introduces a normative parameter into the causal model framework, most commonly precisified as a distinction between default and deviant states of a system (Beckers and Vennekens 2018; Gallow 2021; Hall 2007; Halpern 2016b; Halpern and Hitchcock 2010, 2015; Livengood 2013; Menzies 2017). There are a few reasons to resist this move, however. The first is that the question of when an event is default vs. deviant is a highly complex issue. This stems in part from the fact that there are many different, possibly competing, sets of norms. Halpern and Hitchcock recognize this when they write,

> The problem [of justifying the model] is exacerbated by the fact that default and 'normality' have a number of interpretations. Among other things, they can represent moral obligations, societal conventions, protoypicality information, and statistical information. All of these interpretations are relevant to understanding causality; … (2010:386)

A complete analysis would need to provide principles that say which normative considerations go towards determining whether an event is default or deviant, as well as how competing considerations should be weighed when they deliver contrary verdicts about an event's status. Second, insofar as norms pertaining to domains such as morality or societal conventions are part of what's considered, this move threatens to undermine realism about actual causation.[14] If pragmatically determined norms such as who is allowed to borrow pens from the department office go toward determining what causes what, then what causes what is not a mind- and language-independent feature of reality.[15] Finally, it can be argued that including a normative parameter in our metaphysics of causation in fact leads to a psychologically implausible analysis of actual causation (Blanchard and Schaffer 2017).

It would be nice, then, to have an alternative response to that of introducing a normative parameter. But there is only one such response in the literature – that presented by Blanchard and Schaffer (2017) – and it is incomplete. They argue that the model on the respective interpretation can be ruled inapt for representing **Bogus Prevention** by endorsing a requirement on aptness that I will call *Essential Structure*. This requires of an

---

[14] This consequence – anti-realism about actual causation – is explicitly acknowledged by (Halpern 2016b; Halpern and Hitchcock 2010; Hitchcock 2007a). It is also implied by the view in (Menzies 2004b) insofar as at least some kinds of systems are governed at least in part by human-dependent norms – although (Menzies 2007) suggests that such systems would not be legitimate. Hitchcock, at least, preserves realism at the level of token causal structure.

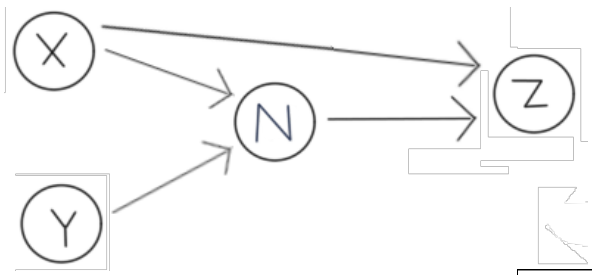
[15] See (Knobe and Fraser 2008) for an example of such a norm affecting causal judgments. For further examples of the general phenomenon see also (Cushman, Knobe, and Sinnott-Armstrong 2008).

apt model-interpretation pair that it represent enough of a situation so as to capture its essential structure.[16] Invoking this principle, $\mathcal{M}_1$ on $\mathcal{I}(\mathcal{M}_1)_{BP}$ is inapt for representing **Bogus Prevention** because it leaves out essential structure – namely, the bit where the antidote neutralizes the poison (or not).[17] Including a variable to represent this produces a model, $\mathcal{M}_{1+}$, that delivers the intuitively correct causal verdict.

$\boldsymbol{\mathcal{S}}$ = $\boldsymbol{U} = \{X, Y\}$
$\boldsymbol{V} = \{N, Z\}$
$\boldsymbol{R} = f\,(X_i) = \{1, 0\}$

$\boldsymbol{\mathcal{A}}$ = (EQ1) $X = 1$
(EQ2) $Y = 1$

$\boldsymbol{\mathcal{L}}$ = (EQ3) $N := \min(1 - X, Y)$
(EQ4) $Z := \max(X, N)$



$\mathcal{M}_{1+}$

$\mathcal{I}(\mathcal{M}_{1+})_{BP}$ is the same as $\mathcal{I}(\mathcal{M}_1)_{BP}$, with *N = 1* representing poison being neutralized in the coffee and *N = 0* representing the coffee not undergoing poison neutralization. According to the new model-interpretation pair, <$\mathcal{M}_{1+}, \mathcal{I}(\mathcal{M}_{1+})_{BP}$>, Bodyguard's administration of antidote is no longer judged an actual cause of King surviving. To see why, though, requires specifying a recipe for reading facts of actual causation off a model. For my purposes, I will adopt that put forward by Blanchard and Schaffer (2017:179), which comes from Hitchcock (2001b:290): [18]

[16] Essential Structure is originally introduced in (Hitchcock 2007a).

[17] That the enriched model delivers the right verdict is noted in (Halpern and Hitchcock 2015), and adding a variable like *N* in this example is also mentioned by Weslake (2015), who attributes the idea to Hitchcock.

[18] See (Gallow 2021; Halpern and Pearl 2005; Hitchcock 2001b; Weslake 2015; Woodward 2003) for various recipes of actual causation.

($\mathbf{AC}$) c is an actual cause of e just in case there is an apt model, $\mathcal{M}_i$, representing c as $X = x$ and e as $Y = y$, such that … [19]

AC1) $X = x$ and $Y = y$ in $\mathcal{M}_i$.

AC2) There is a route P_i in $\mathcal{M}_i$ from X to Y and a possible assignment of values to the set of variables off P_i such that the following counterfactuals are true:[20]

(a) Had off-route variables (call these $\vec{Z}$) taken the specified assignment (call this $\vec{z}$), the variables on P_i would still have taken their actual values.

(b) Had $\vec{Z} = \vec{z}$ and $X = x$, then $Y = y$.

(c) Had $\vec{Z} = \vec{z}$ and $X = x_i$, where $x_i \neq x$, then $Y = y_i$, where $y_i \neq y$.

AC1 is an actuality condition. It requires that the model represent that c and e are actual events. AC2 handles redundant causation and so opens the way for c to actually cause e

[19] This analysis of actual causation is made general by existentially quantifying over apt model-interpretation pairs, which is the standard way to go (Blanchard and Schaffer 2017; Hitchcock 2001b; Weslake 2015). But one could *universally* quantify over apt model-interpretation pairs (Hall 2007) – or indeed quantify in any number of ways. Alternatively, one could provide a model-relative analysis (Halpern and Hitchcock 2015; Halpern and Pearl 2005).

[20] A *route* in a SEM is a sequence of variables, $\langle X_1, X_2, X_3, \ldots, X_i \rangle$, such that X_1 figures on the right-hand side of the X_2-equation, X_2 figures on the right-hand side of the X_3-equation, …, and X_{i-1} figures on the right-hand side of the X_i-equation. The sequence of nodes corresponding to such variables in a corresponding DAG are called a *directed path*, hereafter just *path*, and are such that the arrows between them all point in the same direction.

even when the latter doesn't depend counterfactually on the former (as, for example, with Suzy's throw and the window shattering). It says that there must be a route between the putative cause and effect such that when all off-route variables are held fixed at some permissible value, then intervening to set the putative cause as occurring will result in the effect occurring (AC2b), and intervening to set some alternative to the putative cause will result in some alternative to the effect occurring (AC2c). It also places a condition on permissible values of off-route variables that they preserve the actual values of the on-route variables (AC2a).

With this recipe on the table, we can see how the introduction of *N* renders *Y = 1* a non-cause of *Z = 1* in $\mathcal{M}_{1+}$. The only possible route between *Y* and *Z* is {*Y, N, Z*}, and there is no setting of values of off-route variables, {*X*}, that satisfies AC2a – c. When *X = 1*, AC2c isn't satisfied, and when *X = 0*, AC2a isn't satisfied. Thus, *Y = 1* is not an actual cause of *Z = 1* relative to $\mathcal{M}_{1+}$.

§3 Evident Mediation

However, Essential Structure is inadequate as an objective condition on apt causal models. Without further specification, Essential Structure remains opaque, reliant on our pre-theoretic causal intuition, and unilluminating of the nature of causation. We need an independent story about what kind of structure is essential and why.[21] In this section, I

[21] In fairness, Blanchard and Schaffer themselves concede that this doesn't yet get to the bottom of things. They write, "[W]e think that there is a core phenomenon of *an impoverished model that omits crucial information*. We think that there needs to be some constraint corresponding to the vague idea of 'don't use

argue that the needed constraint is a requirement of *Evident Mediation*. On this view, a model is impoverished and therefore inapt insofar as it omits *partially mediating variables*, which I define shortly, and enriching the model in the relevant sense is just to include them. I defend this proposal by illustrating what work it can do,[22] and showing how the need for this requirement on models is independently motivated.

3.1 More Than One Way to Mediate

Myriad events occur between any given cause and effect. My pulling the trigger causes the gun to fire. Between these two events, the hammer drops. Say an event, *p*, *intercedes* between two events, *x* and *y*, (where *x* and *y* are distinct from *p* and from each other), when *p* depends on *x*, and *y* depends on *p*. The hammer dropping is an *interceding event* in the chain of dependence that begins with my pulling the trigger and ends with the gun firing. When such an event is represented by a variable in a model, it is a *mediating variable*. But there are, in fact, different kinds of mediating variables. I will argue that the difference between variables that *fully* mediate between their flanking variables and those that only *partially* mediate is of special significance.

Consider first fully mediating variables. Assume a scenario similar to **Overdetermination**, but where Billy doesn't throw anything, nor would he have. Suzy

impoverished models'. Any such constraint should equally be able to do the work we put [Essential Structure] … towards. (2017:13)"

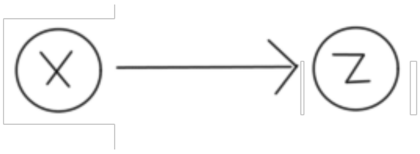
[22] To be clear, Evident Mediation does all the work of Essential Structure but not *all* the work done in (Blanchard and Schaffer 2017). In particular, Blanchard and Schaffer employ a different aptness principle to resolve problems of causation by omission, with which Evident Mediation is not meant to help.

and only Suzy throws a rock and the window shatters. We can represent this with the following model, $\mathcal{M}_2$, where $\mathcal{I}(\mathcal{M}_2)$ is just the same as $\mathcal{I}(\mathcal{M}_1)_O$, above, save the omission of any assignment of content to Y.

$\boldsymbol{\mathcal{S}}$ = $\boldsymbol{U} = \{X\}$
$\boldsymbol{V} = \{Z\}$
$\boldsymbol{R} = f\,(X_i) = \{1, 0\}$

$\boldsymbol{\mathcal{A}}$ = (EQ1) $X = 1$

$\boldsymbol{\mathcal{L}}$ = (EQ2) $Z := X$



$\mathcal{M}_2$

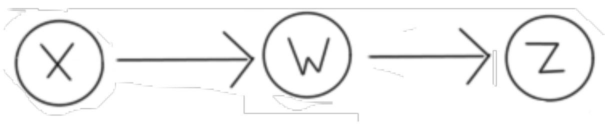
Of the countless events between Suzy throwing her rock and the window shattering, consider that of Suzy's rock hitting the window. The window shattering depends on the rock hitting it, which in turn depends on Suzy's throw. Understood as *counterfactual* dependence, the window shattering counterfactually depends on the rock hitting it, which in turn counterfactually depends on Suzy's throw. Understood as *type-level causal* dependence, events of the window-shattering type causally depend on events of the rock-hitting type, which in turn causally depend on events of the rock-throwing type. So, the rock hitting the window *intercedes* between these two events. Further, given the event of the rock hitting the window, the window shatters *regardless of* Suzy's throwing or not. Counterfactually, had Suzy's rock hit the window even though Suzy had not thrown, then the window still would have shattered. But had the rock failed to hit the window despite Suzy throwing it, the window would not have shattered. No counterfactual dependence of the window shattering on Suzy's throw exists independently of the rock hitting the window. Analogously, in situations of this kind, there's no type-level causal dependence between rock-throws and window-shatterings that is independent of rock-hittings. Thus, there is a simple chain of dependence that holds between the three events.

$\mathcal{M}_2$ on $\mathcal{I}(\mathcal{M}_2)$ doesn't explicitly represent this interceding event. Instead, whenever the model sets X = 1, it implicitly represents the rock hitting the window. And whenever the model sets X = 0, it implicitly represents the rock not hitting. But we *could* introduce a variable to explicitly capture this event. Say we introduce W: {1, 0}, and add to $\mathcal{I}(\mathcal{M}_2)$ the interpretive assignment of W = 1 as the rock's hitting the window and W = 0 as the rock's not hitting. This produces the following amended model, $\mathcal{M}_{2+}$:

$\boldsymbol{\mathcal{S}}$ = $\boldsymbol{U}$ = {X}
$\boldsymbol{V}$ = {W, Z}
$\boldsymbol{R}$ = $f(X_i)$ = {1, 0}

$\boldsymbol{\mathcal{A}}$ = (EQ1) X = 1

$\boldsymbol{\mathcal{L}}$ = (EQ2) $W := X$
(EQ3) $Z := W$



$\mathcal{M}_{2+}$

Upon introducing W, the only remaining route between X and Z in the DAG is through W. This is because once W is held fixed at a value, no variation on X will result in any variation in Z. In the SEM, *W replaces X* in the Z-equation. W cutting X off from Z in the DAG and replacing X in the Z-equation represents the fact that any dependence that the window shattering (or not) has upon Suzy's throwing (or not) it has *completely in virtue of* the event of the rock hitting (or not). Say that a variable, W, is a *mediating variable* between X and Z just in case X figures in the W equation – i.e. X is a parent of W – and W figures in the Z equation – i.e. W is a parent of Z. Then, a variable, W, is *fully mediating* between X and Z just in case W mediates between X and Z, and W replaces X in the Z-equation – i.e.

W is a parent of Z and X is no longer a parent of Z.[23] We can now see that an interceding event, as defined above, will always be represented by a fully mediating variable.

A variable that mediates but doesn't fully mediate is merely a *partially mediating* variable. A helpful illustration of partial mediation comes from **Bogus Prevention**. Consider the actual event of the coffee not undergoing poison neutralization. This event does not intercede between Assassin not administering poison and King surviving, since King surviving doesn't depend (in these circumstances, at least, where there's no poison) on the coffee not undergoing poison neutralization. Analogously, survivals don't causally depend on neutralizations in situations like our target one, in which there's no poison. A variable representing this event therefore wouldn't fully mediate between the variables representing King's surviving and Assassin not administering poison. And yet, the event of the coffee not undergoing poison neutralization *can* be represented by a mediating variable, as it is in $<\mathcal{M}_{1+}, \mathcal{I}(\mathcal{M}_{1+})_{BP}>$. While there isn't a simple chain of dependence here, there is different kind of dependence. On the one side, the coffee being neutralized (or not) simply depends on Assassin not administering poison. Poison isn't administered and neutralization doesn't occur, but had poison been administered, neutralization would have occurred. Analogously, neutralizations causally depend on poisonings in situations,

[23] This definition is overly simple in that the notion is somewhat broader. The comprehensive definition also covers cases where a variable, W, fully mediates between *sets* of variables, with more than one member on one or both sides. Suppose there are two sets of variables, $\boldsymbol{X}$ and $\boldsymbol{Z}$. Then, when a mediating variable, W, is introduced between $\boldsymbol{X}$ and $\boldsymbol{Z}$, W *fully mediates* between them just in case (i) every $X : X \in \boldsymbol{X}$ figures in the W-equation, and (ii) W replaces each $X : X \in \boldsymbol{X}$ in every Z-equation where $Z \in \boldsymbol{Z}$. This coincides with the definition in the main text whenever each set, $\boldsymbol{X}$ and $\boldsymbol{Z}$, has only one member.

like our target one, in which there's antidote. So, the variable for the poison administration – call it *X* – figures in the equation for the neutralization variable, *N*.

But on the other side, things aren't so simple. We don't have simple dependence between King's surviving and the coffee not being neutralized, nor do we have simple dependence between King's surviving and Assassin not administering poison. But notice that King's survival *does* depend on whether Assassin administers poison *given* the coffee doesn't undergo neutralization. Thus, whether the King survives depends on each of whether neutralization occurs and whether poison is administered only under a certain contingency about the other. Call this *de facto dependence*. I will say more about the nature of this dependence in the next section. For now, notice that were we to represent all this in a model, variables representing both neutralization occurrence and poison administration would need to figure in the equation for the variable representing King's survival.

Indeed, this is what we see in $<\mathcal{M}_{1+}, \mathcal{I}(\mathcal{M}_{1+})_{BP}>$. As a representation of **Bogus Prevention**, $\mathcal{M}_1$ doesn't explicitly represent the additional event of the coffee being neutralized. But it could. Doing so results in the introduction of *N*: {1, 0}, and we get $<\mathcal{M}_{1+}, \mathcal{I}(\mathcal{M}_{1+})_{BP}>$. In the corresponding DAG, there is a route from *X* to *Z* that goes through *N*. But there remains an *independent* route between *X* and *Z*, as well – one that does not pass through *N*. This is because it is not the case that, so long as *N* is held fixed, there is no variation on *X* that would result in variation in *Z*. There is at least one value of *N* (namely, *N =0)* such that, holding *N* fixed at that value, variation in *X still* results in variation in *Z*. In the SEM, *N* does not *replace X* in the *Z*-equation, though it does make its own appearance. Call a variable like *N* a *partially mediating* variable. In general, when a mediating variable, N, is introduced between two variables, X and Z, the introduced variable *partially mediates* just in case both X and N figure in the Z-equation.[24]

3.2 Evident Mediation

The difference matters. Omission of a fully mediating variable is benign – a model-interpretation pair that omits such a variable may still aptly capture the causal structure of a situation (assuming it meets the conditions for accuracy). The omission of uncountably many fully mediating variables is, in fact, formally necessitated by the finite nature of an SEM coupled with the presumably dense nature of reality. However, omission of a partially mediating variable will make for an inapt model-interpretation pair. The inaptness of $\mathcal{M}_1$ for representing **Bogus Prevention** is due to its leaving *implicit* a partially mediating variable. Rendering this explicit is what makes $\mathcal{M}_{1+}$ apt. Why? Partially mediating variables capture a particular kind of dependence – *de facto dependence* – that is both highly relevant for causation and distinctly different from simple counterfactual dependence.[25] Dependence between two events, c and e, is *simple* (or holds *simpliciter*) when e depends on c full stop. In contrast, dependence between c and e is *de facto* (or *de facto* holds) when e depends on c *only* when certain other features – call these *anchoring event(s)* – are held fixed, where these other features themselves

[24] Again, this definition is overly simple for the same reason that the notion is somewhat broader. A variable, W, may partially mediate between *sets* of variables with more than one member on one or both sides. Expansion to the comprehensive definition follows the same lines as the expansion found in fn. 23.

[25] This includes the kind of dependence captured by what Hitchcock calls *Explicitly Non-Foretracking (ENF) counterfactuals*. See especially (Hitchcock 2001b). See also (Yablo 2002), although his account diverges from that offered here.

depend simpliciter on *c*.[26] An anchoring event, as just defined, will always be represented by a partially mediating variable. The need for explicit representation of de facto dependence should be clear. As was illustrated in §2.3, SEM recipes involve first singling out a route between two variables and then discriminating between variables that are on-route versus off-route, holding off-route variables fixed and allowing on-route variable to vary in accord with their equations. This ability of SEMs is so useful precisely because it means they have the resources to capture both simple *and* de facto dependence, incorporating both into an analysis of causation. This is what allows for their trademark solutions to redundant causation, which I discuss further in §4.1. It is unsurprising that the success of such an analysis will be compromised when the model-interpretation pair represents what is in fact de facto dependence by a single independent route, omitting representation of the crucial anchoring event.[27] In line with this, I propose the following aptness requirement:

Evident Mediation (EM) For any two variables in a model, *X* and *Z*, every event, *p*, is explicitly represented in the model – where *p* is such that were it represented in

[26] Note that, according to this definition, no mere background condition will qualify as an anchoring event – since they will not themselves depend on the putative cause event. Thanks to an anonymous referee for helping clarify this point.

[27] Not all SEM analyses of actual causation are in terms of routes. Some are in terms of variable sets. See for example: (Halpern 2016a; Halpern and Pearl 2005). Arguably, however, such analyses work because the distinction between on-route and off-route variables can be put in terms of distinct sets of variables, once again providing the resources to capture de facto dependence. Since the sequential order of a route is fixed by the equations, routes map one-to-one onto variable sets so long as the equations are fixed.

the model, it would be represented by a partially mediating variable between X and Z and the resulting model would still be accurate.[28]

Since a partially mediating variable always represents an anchoring event, another way to put this is that whenever a model represents de facto dependence between two events, it must also explicitly represent the anchoring event relative to which that dependence holds. A model-interpretation pair will be apt only if EM is satisfied.

The spirit of Evident Mediation can be found in (Gallow 2021). Gallow similarly emphasizes the need to be clear about whether a variable determines another along a single route or along multiple routes. He proposes a necessary principle of aptness whereby *if* a model is apt, then the removal of an *inessential* variable will still produce an apt model. The principal reason for an endogenous variable's being inessential, on his view, is its being fully mediating.[29] Gallow agrees, then, that fully mediating variables may be benignly omitted. Further, Gallow can be seen as considering partially mediating variables essential – insofar as partially mediating variables are the only other kind of variable. However, his discussion focuses only on conditional principles of aptness. He is

[28]This proposal can be seen as a way of rendering concrete proposals by Hitchcock (2001a:fn21) and by Hiddleston (2005a:649). It is also one way of precisifying an idea of Halpern and Hitchcock, who propose that only when the addition of variables to a model changes its "topology" will those additions affect the relations of actual causation (Halpern and Hitchcock 2010:395), under the assumption that the introduction of a partially mediating variable is the only topologically relevant change.

[29] I'm using my own terminology here. Gallow uses the term "correctness" to refer to what I call aptness and "interpolating variable" to refer to what I call a fully mediating variable. He doesn't introduce a term to refer to what I call partially mediating variables.

silent about how to check for aptness were we to *start* with a model that has omitted a partially mediating variable. Presumably, though, he would want to say that such a model would be inapt, and that it would be inapt precisely because it omits a partially mediating variable. After all, the only difference is where we start! So, according to both EM and to what I am assuming must be Gallow's view, if a model omits a partially mediating variable, it is thereby inapt. Gallow goes on to argue for the inclusion of a normative parameter, as well, with which to solve the problem of structural isomorphs. I argue that Evident Mediation is all we need.

In further support of Evident Mediation, it's worth pointing out how it provides a general answer to Hall's (2007) original concern which launched the discussion of structural isomorphs. Hall argues that a SEM analysis of actual causation will always mistakenly ascribe causation to any preventative measure regardless of whether it actively protects against a live threat or merely safeguards against possible but non-actual threats. This is the difference between the neighborhood patrol stopping a burglary from taking place, and so causing the family to sleep peacefully through the night, and the neighborhood patrol merely safeguarding the family from any possible burglaries, although none are actually attempted. In the first instance the neighborhood patrol causes the family to sleep peacefully through the night by actively preventing the burglary. In the second instance, the neighborhood patrol does not cause the family to sleep peacefully, although it would have had a burglary been attempted. Note that Bodyguard's administration of antidote is just such a safeguard, too, in a situation without poison such as **Bogus Prevention**.

Take it as given, then, that there is a real causal difference between active protectors and mere safeguards. In order to distinguish them, Hall insists that we need a default/deviant distinction. Alternatively, proponents of Essential Structure can claim that some essential structure must be missing – though they lack a principled account of "essential". Evident Mediation fills this in. For any model that accurately represents a situation where the preventative measure is merely a safeguard, and which explicitly represents the threat against which the safeguard protects, there may always be introduced a variable to represent whether the prevention – whatever it is – occurs or not. The introduced variable will partially mediate between the (non-existent) threat and the safeguard, on the one side, and the effect in question, on the other. Once it is explicitly included, the safeguard no longer satisfies **AC** as a cause of the effect in question, while the protector would. Evident Mediation gives us a way to distinguish between active protectors and mere safeguards without relying on a normative parameter.

§4 Further Application and Objections

There is good reason already, then, to adopt this proposal. But before engaging with further objections one might have, I will discuss two other reasons in support of its adoption: that Evident Mediation predicts and justifies the standard solution to the problem of redundant causation, and that it can do the work done by Essential Structure to illuminate another case of problematic structural isomorphism in the literature – **Bogus Antidote**. I should flag that while my own view is that EM solves everything that goes wrong with **Bogus Antidote**, one may have conflicting intuitions. If so, discussion of the case is included only to show once more that the "essential structure" introduced by Blanchard and Schaffer is no different from that captured by *partially mediating variables*.

4.1 Redundant Causation

Redundant causation refers to cases in which counterfactual dependence between a cause and an effect fails to hold due to the presence of a backup cause waiting in the wings. This includes late and early preemption. Consider *late* preemption first:

> **Late Preemption** Suzy and Billy each throws a rock at a window. Suzy's rock hits the window first and the window shatters. Moments later, Billy's rock sails through the air where the window used to be.

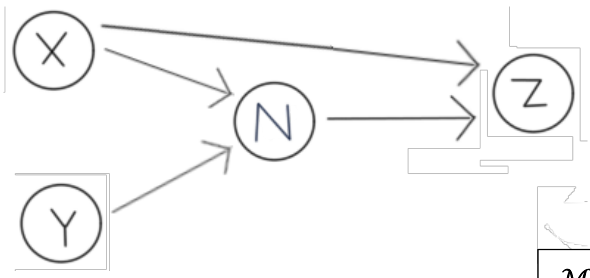
Suzy's throw causes the window to shatter, despite the fact that had Suzy's rock not hit, then Billy's would have, and the window still would have shattered. Notice that we *could* model this with $\mathcal{M}_1$ from above, indeed even using the same interpretation as **Overdetermination**! But when **AC** is applied to $\mathcal{M}_1$, it says that both *X = 1 and Y = 1* are actual causes of *Z = 1*. This is not the result we want for **Late Preemption**. *Y = 1* represents Billy's throw, which is obviously *not* a cause of the window shattering. What's gone wrong? The standard SEM solution to this problem introduces a variable that represents Billy's rock hitting the window.[30] This results in the following:

[30] See, for example, (Halpern and Pearl 2005; Menzies 2004a; Menzies and Beebee 2019). In fact, the standard solution also introduces a variable to represent *Suzy's* rock hitting the window, but this isn't strictly needed. Further, EM doesn't require it. Such a variable would *fully* mediate between, on the one side, the variable representing Suzy's throw and, on the other, the variables representing Billy's rock hitting and the window shattering. (In effect, Suzy's rock hitting would take the place of *X*, with Suzy's throw represented by a new exogenous variable upstream.)

$\mathcal{S}$ = $\boldsymbol{U}$ = {X, Y}
$\boldsymbol{V}$ = {N, Z}
$\boldsymbol{R}$ = f (X_i) = {1, 0}

$\mathcal{A}$ = (EQ1) $X = 1$
(EQ2) $Y = 1$

$\mathcal{L}$ = (EQ3) $N := \min(1 - X, Y)$
(EQ4) $Z := \max(X, N)$



$\mathcal{M}_{1+}$

$\mathcal{I}(\mathcal{M}_{1+})_{LP}$ is the same as $\mathcal{I}(\mathcal{M}_1)_O$, with *N = 1* representing Billy's rock hitting the window and *N = 0* representing it not hitting. According to $\mathcal{M}_{1+}$, only *X = 1* is a cause of *Z = 1*. This point should be familiar from earlier. The only possible route between *Y* and *Z* is {*Y, N, Z*}, and there is no setting of values of off-route variables, {*X*}, that satisfies **AC**.

While introducing extra variables resolves late preemption, the move may seem ad hoc. Evident Mediation provides independent motivation. EM deems $\mathcal{M}_1$ an inapt model for representing **Late Preemption** due to its omission of a partially mediating variable and calls for the introduction of *N*. It's a happy consequence that once this *N* is included, the model-interpretation pair delivers the result we want: Suzy's throw is and Billy's throw is not a cause of the window shattering.

Next, consider a case of *early* preemption:

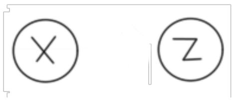
Early Preemption Suzy throws a rock at a window, the rock hits the window, and the window shatters. Her friend Billy stands by. Had Suzy not thrown, then Billy would have. And had Billy thrown, the window would still have shattered.

Now, we *could* represent this with the following model, interpreted in the same way:

$\boldsymbol{\mathcal{S}}$ = $\quad$ $\boldsymbol{U}$ = {*X*, *Z*}
$\boldsymbol{V}$ = {∅}
$\boldsymbol{R}$ = *f* (*X*$_i$) = {1, 0}

$\boldsymbol{\mathcal{A}}$ = $\quad$ (EQ1) $X = 1$
(EQ2) $Z := 1$

$\boldsymbol{\mathcal{L}}$ = $\quad$ ∅



$\mathcal{M}_3$

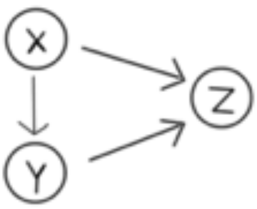
$$\mathcal{I}(\mathcal{M}_3)_{EP}: \qquad X = \begin{cases} 1 \; if \; Suzy \; throws \; a \; rock \\ 0 \; if \; Suzy \; doesn't \; throw \end{cases} \qquad Z = \begin{cases} 1 \; if \; the \; window \; shatters \\ 0 \; if \; the \; window \; doesn't \; shatter \end{cases}$$

But $\mathcal{M}_3$ fails to capture the causal relation between Suzy's throw and the window shattering. Pre-theoretically, we recognize that the dependence between them is conditional on Billy *not* throwing. The natural move, then, is to include Billy's throw in the model. In fact, Evident Mediation gives us a principled reason for this: EM requires the introduction of a variable to represent Billy's throw, since it's a partially mediating variable. <$\mathcal{M}_3$, $\mathcal{I}(\mathcal{M}_3)_{EP}$> is inapt for representing **Early Preemption**. Instead, the enriched model-interpretation pair, <$\mathcal{M}_{3+}$, $\mathcal{I}(\mathcal{M}_{3+})_{EP}$>, is apt.

$\boldsymbol{\mathcal{S}}$ = $\quad$ $\boldsymbol{U}$ = {*X*}
$\boldsymbol{V}$ = {*Y*, *Z*}
$\boldsymbol{R}$ = *f* (*X*$_i$) = {1, 0}

$\boldsymbol{\mathcal{A}}$ = $\quad$ (EQ1) $X = 1$

$\boldsymbol{\mathcal{L}}$ = $\quad$ (EQ2) $Y := (1 - X)$
(EQ3) $Z := \max(X, Y)$



$\mathcal{M}_{3+}$

$\mathcal{I}(\mathcal{M}_{3+})_{EP}$ is the same as $\mathcal{I}(\mathcal{M}_3)_{EP}$, with *Y = 1* representing Billy throwing a rock and *Y = 0* representing Billy not throwing a rock. According to $\mathcal{M}_{3+}$, *X = 1* is a cause of *Z = 1* when *Y* is held fixed at its actual value of *Y = 0*. $\mathcal{I}(\mathcal{M}_{3+})_{EP}$ interprets this as Suzy's throw causing the window to shatter. EM motivates and justifies the move to this enriched model, which reveals the causal relation we know to exist.

4.2 Short Circuits

Next, consider the following instance of what Hall calls a "short circuit" (2007:120):

> **Bogus Antidote** Bodyguard accidentally spills some antidote into King's coffee. Assassin sees this. She has an obligation to poison the coffee, but doesn't want to actually kill King. Now, she can poison the coffee without risking killing King. She does so. King drinks the coffee and survives. [31]

Intuitively, there is *something off* about saying that the bodyguard's putting the antidote in the coffee is a cause of King surviving. Precisely what, though, is a point of controversy. It is true that the antidote prevents the poison from killing King, but the only reason for there being poison in the first place is the presence of the antidote. The only threat subdued by the antidote is one that it produces! This motivates some to argue that Bodyguard's administration of antidote is not an actual cause of King's survival (Blanchard and Schaffer 2017; Gallow 2021; Hitchcock 2007a; Weslake 2015). The problem, however, is that **Bogus Antidote** is structurally isomorphic to **Early**

[31] This example comes directly from (Blanchard and Schaffer 2017:202), and goes by the name "Careful Poisoning" in (Beckers and Vennekens 2018:12; Weslake 2015). It is originally given under the name "Counterexample to Hitchcock" by Hitchcock (2007a:519), who attributes similar, albeit messier, examples to Michael McDermott (personal communication), (Bjornsson 2006), and (Hitchcock 2003).

Preemption. The same model we just saw, $\mathcal{M}_{3+}$, can be interpreted so as to accurately represent both situations. Use the same interpretation as above, $\mathcal{I}(\mathcal{M}_{3+})_{EP}$, to represent **Early Preemption**, and use a new interpretation, $\mathcal{I}(\mathcal{M}_{3+})_{BA}$, to represent **Bogus Antidote**:

$$\mathcal{I}(\mathcal{M}_{3+})_{BA}: \quad X := \begin{cases} 1 \; if \; Bodyguard \; administers \; antidote \\ 0 \; if \; Bodyguard \; doesn't \; administer \; antidote \end{cases}$$

$$Y := \begin{cases} 1 \; if \; Assassin \; doesn't \; administer \; poison \\ 0 \; if \; Assassin \; administers \; poison \end{cases} \qquad Z := \begin{cases} 1 \; if \; King \; survives \\ 0 \; if \; King \; dies \end{cases}$$

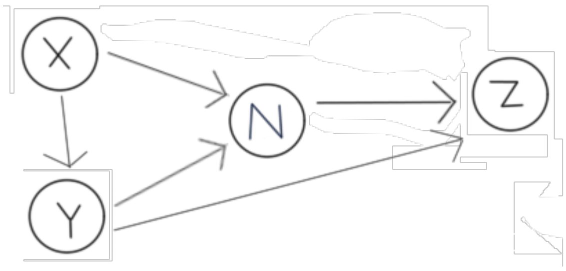
When **AC** is applied to $\mathcal{M}_{3+}$, it says that *X = 1* is an actual cause of *Z = 1*. The route, {*X*, *Z*} is such that when off-route variables (*Y*) are held fixed at their actual values (*Y = 0*), then if it were the case that *X = 1* then *Z = 1*, and if it were the case that *X = 0* then *Z = 0*. This was the result we wanted for **Early Preemption**. *X = 1* represents Suzy's throw, which is indeed the intuitive cause of the window shattering. However, this is not the result we want for **Bogus Antidote**. *X = 1* represents Bodyguard's administration of antidote, which, again, is supposed to *not* be an actual cause of King surviving.

Blanchard and Schaffer (2017) use Essential Structure to respond to this problem. They argue that $\mathcal{M}_{3+}$ is impoverished in leaving out essential structure, namely whether there is neutralization of poison or not. $\mathcal{M}_{3+}$ is ruled inapt for representing **Bogus Antidote** and the isomorphism between **Bogus Antidote** and **Early Preemption** is broken. Observe that the omitted essential structure is a partially mediating variable. Evident Mediation therefore also requires explicit representation of this feature – the neutralization – and similarly breaks the isomorphism. Add variable *N,* with *N = 1* representing neutralization of poison and $N = 0$ representing no neutralization, and call the resulting model $\mathcal{M}_{3++}$:

$\boldsymbol{\mathcal{S}}$ = $\boldsymbol{U} = \{X\}$
$\boldsymbol{V} = \{Y, N, Z\}$
$\boldsymbol{R} = f(X_i) = \{1, 0\}$

$\boldsymbol{\mathcal{A}}$ = (EQ1) $X = 1$

$\boldsymbol{\mathcal{L}}$ = (EQ2) $Y := (1 - X)$
(EQ3) $N := \min(X, (1 - Y))$
(EQ4) $Z := \max(N, Y)$



$\mathcal{M}_{3++}$

Still, even relative to $\mathcal{M}_{3++}$, Bodyguard's administration of antidote remains a cause of King surviving. That is, $X = 1$ is still a cause of $Z = 1$. Insofar as this is the wrong result, enriching the model didn't solve the problem. Then, what purpose did enriching the model serve? Blanchard and Schaffer justify the enrichment as permitting us to now represent how the administration of antidote might have led to the King's death. Had it somehow failed to nullify the poison, then the antidote would have caused the King's death. This possibility, they argue, is why it seems counterintuitive to say that the antidote caused the King's survival. And this possibility is only represented in the enriched model (by setting $N = 0$). As to **AC** still delivering the wrong result in the enriched model, Blanchard and Schaffer suggest this may simply indicate **AC**'s inadequacy. And now that the structural isomorphism is broken, the right amendment of **AC** may deliver the correct result in both **Early Preemption** and **Bogus Antidote**. In other words, what's left is not a problem of aptness but a problem for the recipe of actual causation to solve.

Of course, an alternative response is that there's nothing left to solve: Bodyguard's administration of antidote *is* an actual cause of King surviving.[32] After all, in order for EQ3 of $\mathcal{M}_{3++}$ to be true of **Bogus Antidote**, the amount and kind of antidote administered by Bodyguard *must* nullify whatever poison Assassin administers. Thus, while the model might seem to say simply that Bodyguard administering antidote is a cause of King's survival, what it actually says is that Bodyguard administering *enough* of the *right kind* of antidote such that it nullifies Assassin's poison is a cause of the King's survival. Intuitively, this doesn't seem so wrong.

Further, and as Blanchard and Schaffer point out, our resistance to the causal claim that the administration of antidote causes King to survive perhaps stems from the role causal judgments play in responsibility ascriptions. Such a claim on its own would normally result in approval for Bodyguard. However, the full story includes the fact that the only threat to King's life protected against by Bodyguard's action was also caused by it! Thus, Bodyguard's good standing would be negated.[33]

4.3 Some Concerns about Evident Mediation

[32] Such a response is in agreement with (Beckers and Vennekens 2018).

[33] Overall, this response to **Bogus Antidote** may strike some as less than satisfying. If one is committed to the *right* result being that the administration of antidote is not a cause, the problem has not yet been solved – it depends on arriving at the right recipe of causation. For those with this commitment, accounts that introduce a normative parameter retain the advantage of being able to deliver the "right" result in these cases. In fact, getting this result on this kind of case seems to be the only reason to prefer Gallow's view (2021) over mine.

One might be concerned that Evident Mediation is too strong or, indeed, impossible to satisfy. It will be worth looking at the reasons one may think this, to see how EM works.

First, one might be concerned that EM demands over-complicating many simple or ordinary SEMs.[34] However, as it turns out, a good number of events that seem like candidates for being required by EM are such that their introduction into a model-interpretation pair would either violate accuracy or else require changing the given interpretation of at least one variable in the model. Notice that Evident Mediation, like any principle of aptness, is satisfied or not by a given model under a given interpretation. This means that whether a model has inappropriately left de facto dependence implicit depends on what variables are already included in the model *and* how those variables have already been interpreted (and, of course, on what situation is being represented). If representation of an event can be introduced into a model as a partially mediating variable *only by* altering the interpretation of other variables in the original model, then the original model-interpretation pair does not violate EM by omitting representation of the event in question. For any such an event, EM does not call for its introduction.

To see this in action, say we model a situation just like **Bogus Prevention** but where Bodyguard doesn't actually administer any antidote. We only include two variables: one to represent Assassin not administering poison and one for King surviving. Such a model would *not* omit any partially mediating variables. This may be surprising. One might think that it needs to include a variable for neutralization or not. Didn't we just get through saying that such a variable must be included for **Bogus Prevention**? But remember that

[34] Thanks to an anonymous referee for highlighting this concern.

we've supposed Bodyguard has *not* administered any antidote. Given this fact, the event of the coffee not undergoing poison neutralization does not counterfactually depend on Assassin's administration of poison. Had Assassin administered poison or had Assassin *not* administered poison, there would be no neutralization. So, were it represented by a variable in the model, neutralization wouldn't be represented as a partially mediating variable. If we were to first introduce a variable to represent whether Bodyguard administers antidote, *then* Evident Mediation would require a variable to represent the event of there being no neutralization. At that point, the model would be capable of representing the possible situation in which antidote is administered. And under that supposition, King surviving would de facto depend on Assassin administering poison, with the anchoring event being whether neutralization occurs. That is, introducing a variable for Bodyguard administering antidote introduces into the model a representation of a possible de facto dependence between King surviving and Assassin administering poison, in addition to the dependence simpliciter between them that holds in the actual situation. We therefore need to ensure the anchoring event relative to which this possible de facto dependence would hold is also represented. EM does this work.

A different reason for thinking Evident Mediation is too strong is that it may seem like any finite model must omit at least some partially mediating variables. To see this, say an actual cause, *c*, and its actual effect, *e*, are respectively represented by two variables, *X* and *Z*, in a model. For any such cause-effect pair, <*c*, *e*>, there is at least one event, *p*, such that had *p* occurred then *c* would have failed to cause *e*. The striking of the cue ball, *X = 1*, causes the 8-ball to go into the corner pocket, *Z = 1*, *only under the condition* that it is not blocked (and only under the condition that it does not explode, the condition that it does not dissolve, does not quantum tunnel, teletransport, etc.). But the event of any one of these *not* occurring could be represented by a variable, and that variable would partially mediate between *X* and *Z*.

But this reason doesn't hold up. The *sole* introduction of a variable, *W*, representing such a possible event as the cue ball (not) being blocked would fail to figure in the *X*-equation. In the situation in question, the cue ball not being blocked does not depend on the striking of the cue ball. There is no source of blocking, and so whether or not the cue ball is struck there won't be a blocking. *W* would therefore fail to mediate, let alone partially mediate, between *X* and *Z*. So, EM permits its exclusion. Of course, that *W* fails to partially mediate is contingent on the model not explicitly representing the possible existence of a blocking apparatus. If instead such a variable, *V*, *were* already included for whatever reason in the model, then *W would* partially mediate, and EM would *then* call for including *W*. Once we have *V*, we have the possibility of *V* being set to a blockage being present, in which case whether blocking occurs – the value of *W* – will depend on whether the cue ball is struck – the value of *X*. Thus, *X would* figure in the *W* equation, and *W* would figure in the *Z* equation alongside *X* (that is, without replacement). But if there's no *V*, then no *W* is required by EM. As formulated, EM only calls for introducing a single variable such that, were *that* variable (alone) to be added it would partially mediate.[35]

[35] To clarify, EM may call for the introduction of multiple variables if, when each variable is introduced on its own, it is a partially mediating variable. Further, it may be EM does not call for the introduction of some variable, *W*, on an initial application but *does* call for its introduction after multiple applications. This may occur if *W* would not partially mediate upon its individual introduction, but some other variable, *V*, does partially mediate and so is required by EM. And then, once *V* is included *W now* partially mediates upon its individual introduction. Thanks to an anonymous referee for pushing for this clarification.

This is also what responds to a concern that Evident Mediation is, at least in some cases, *impossible* to satisfy. Suppose in **Early Preemption** (or in **Late Preemption**) that Billy's rock enforces a gravitational pull on Suzy's rock that is required in order for her rock to stay on course to hit the window. Wouldn't EM require we represent this causal structure with partially mediating variables in the model? And supposing time is dense, wouldn't it be impossible to represent this structure in a finite model? As it happens, there is no *single* variable that could be introduced to capture this dependence of the location of Suzy's rock on the location of Billy's, which would partially mediate between any of the three existing variables. We would need to introduce at least two new variables – one for the location of Suzy's rock and one for that of Billy's, for example. Since EM does not call for such an introduction, the gravitational dynamic in play can remain implicit in the model.

Finally, one might be concerned that EM is simply too difficult to work with. Methodologically, it may not be transparent to us when representation of a partially mediating variable has been omitted. How can we ensure that such an event has not escaped our notice? This is a great question that calls for further inquiry. Briefly, I'd say that our way of knowing whether EM is satisfied is partly whether there is an obvious omitted partially mediating variable and partly whether the model-interpretation pair successfully reproduces our pre-theoretic intuitions. If there's no such variable and the model-interpretation pair successfully reproduces our pre-theoretic intuitions, then EM is likely satisfied. If either there is such a variable or the model-interpretation pair doesn't successfully reproduce our pre-theoretic intuitions, then EM is likely not satisfied. Now, one might infer from this reliance on pre-theoretic causal intuition that the resulting analysis of causation is not reductive, after all! This would be a mistake. What makes it the case that the satisfaction conditions for EM hold is distinct from how we might go about checking whether they hold. The former makes no recourse to pre-theoretic causal intuition. The satisfaction conditions are just that there's no event meeting the criteria of partially mediating between any two represented events in the model-interpretation pair. So, invocation of pre-theoretic causal intuition in this way is not a problem for reduction. Is it any other kind of problem? I concede it's not ideal – ideally, we'd have either some independent way to check whether EM is satisfied or else some independent check on our pre-theoretic causal intuition. But this strikes me as an epistemological and practical challenge relevant to the project of *causal discovery*, rather than a metaphysical one that needs answering before we can say we have a complete metaphysical analysis of causation. However, this may be taken to weaken the *normative force* of an analysis of causation – that is, the ability of such an analysis to falsify our causal intuitions.[36] The thought here is something like if the verdicts of the analysis don't line up with intuition, then intuition can simply claim that EM must not be satisfied. Of course, this at the very least shifts the burden of proof. It would then be intuition's responsibility to produce the needed partially mediating variable.

§5 Conclusion

I have argued that sensitivity to the presence (or absence) of partially mediating variables distinguishes between the putatively structurally isomorphic situations in problem cases. Evident Mediation therefore does the work of Essential Structure and can effectively replace it as an aptness condition. Is EM all we need, then, in addition to accuracy? I

[36] Thanks to an anonymous referee for raising this concern.

suspect it is. Another oft-mentioned requirement, Stability, places a condition on an apt model that merely "[a]dding additional variables should not overturn the causal verdicts (Blanchard and Schaffer 2017:183)."[37] It seems likely that Evident Mediation will obviate any additional need for Stability. But the literature on this principle is somewhat disjoint – with some assuming an existentially quantified recipe while others assume a model-relative recipe.[38] I therefore leave full discussion of this further application of EM for another time.

§6 References

Beckers, Sander, and Joost Vennekens. 2017. "The Transitivity and Asymmetry of Actual Causation." *Ergo* 4(1):1–27.

Beckers, Sander, and Joost Vennekens. 2018. "A Principled Approach to Defining Actual Causation." *Synthese* 195(2):835–62.

Bernstein, Sara. 2015. "The Metaphysics of Omissions." *Philosophy Compass* 10(3):208–18.

Bjornsson, Gunnar. 2006. "How Effects Depend on Their Causes, Why Causal Transitivity Fails, and Why We Care About Causation." *Philosophical Studies* 133:349–90.

Blanchard, Thomas, and Jonathan Schaffer. 2017. "Cause Without Default." Pp. 175–214 in *Making a Difference: Essays on the Philosophy of Causation*, edited by H. Beebee, C. Hitchcock, and H. Price. Oxford University Press.

Briggs, Rachael. 2012. "Interventionist Counterfactuals." *Philosophical Studies* 160(1):139–66. doi: 10.1007/s11098-012-9908-5.

Cushman, F., J. Knobe, and W. Sinnott-Armstrong. 2008. "Moral Appraisals Affect Doing/Allowing Judgments." *Cognition* 108(1):281–89.

[37] See also (Halpern 2016b; Halpern and Hitchcock 2010). Hall (2007:12) calls this 'extendability'. Note that this is arguably distinct from the notion of 'stability' discussed by Woodward (Woodward 2006, 2010, 2016, 2018) in the context of causal explanation.

[38] Thanks to an anonymous referee for drawing my attention to this.

Eells, Ellery. 1988. “Probabilistic Causal Interaction and Disjunctive Causal Factors.” Pp. 189–209 in *Probability and Causality*. Vol. 192. Dordrecht: Springer.

Eells, Ellery. 1991. *Probabilistic Causality*. Cambridge: Cambridge University Press.

Gallow, J. Dmitri. 2016. “A Theory of Structural Determination.” *Philosophical Studies* 173(1):159–86.

Gallow, J. Dmitri. 2021. “A Model-Invariant Theory of Causation.” *Philosophical Review*.

Hall, N. 2007. “Structural Equations and Causation.” *Philosophical Studies* 132(1):109–36. doi: 10.1007/s11098-006-9057-9.

Halpern, Joseph. 2000. “Axiomatizing Causal Reasoning.” *Journal of Artificial Intelligence Research* 12:317–37.

Halpern, Joseph, and Christopher Hitchcock. 2010. “Actual Causation and the Art of Modeling.” Pp. 383–406 in *Causality, Probability, and Heuristics: A Tribute to Judea Pearl*. London: College Publications.

Halpern, Joseph Y. 2016a. *Actual Causality*. MIT Press.

Halpern, Joseph Y. 2016b. “Appropriate Causal Models and the Stability of Causation.” *The Review of Symbolic Logic* 9(01):76–102. doi: 10.1017/S1755020315000246.

Halpern, Joseph Y., and Christopher Hitchcock. 2015. “Graded Causation and Defaults.” *The British Journal for the Philosophy of Science* 66(2):413–57.

Halpern, and Judea Pearl. 2005. “Causes and Explanations: A Structural-Model Approach. Part I: Causes.” *The British Journal for the Philosophy of Science* 56(4):843–87. doi: 10.1093/bjps/axi147.

Hendrickson, Noel. 2006. “Towards a More Plausible Exemplification Theory of Events.” *Philosophical Studies* 129:349–75.

Henne, P., A. Pinillos, and F. De Brigard. 2016. “Cause by Omission and Norm: Not Watering Plants.” *Australasian Journal of Philosophy* 95:1–14.

Hiddleston, E. 2005a. “A Causal Theory of Counterfactuals.” *Nous* 39(4):232–57.

Hiddleston, E. 2005b. “Causal Powers.” *British Journal for the Philosophy of Science* 56:27–59.

Hitchcock, Christopher. 2001a. “A Tale of Two Effects.” *Philosophical Review* 110:361–96.

Hitchcock, Christopher. 2001b. “The Intransitivity of Causation Revealed in Equations and Graphs.” *The Journal of Philosophy* 98(6):273–99. doi: 10.2307/2678432.

Hitchcock, Christopher. 2003. “Of Humean Bondage.” *British Journal for the Philosophy of Science* 54:1–25.

Hitchcock, Christopher. 2004. "Routes, Processes, and Chance-Lowering Causes." in *Cause and Chance: Causation in an Indeterministic World*, edited by P. Dowe and P. Noordhof. Routledge.

Hitchcock, Christopher. 2007a. "Prevention, Preemption, and the Principle of Sufficient Reason." *The Philosophical Review* 116(4):495–532.

Hitchcock, Christopher. 2007b. "What's Wrong with Neuron Diagrams?" Pp. 4–69 in *Causation and Explanation*, edited by J. K. Campbell, M. O'Rourke, and H. S. Silverstein. MIT Press.

Hitchcock, Christopher. 2012. "Events and Times: A Case Study in Means-Ends Metaphysics." *Philosophical Studies* 160(1):79–96.

Kim, Jaegwon. 1976. "Events as Property Exemplifications." Pp. 159–77 in *Action Theory*, edited by M. Brand and D. Walton. Dordrecht: Reidel.

Knobe, J., and B. Fraser. 2008. "Causal Judgment and Moral Judgment: Two Experiments." Pp. 441–47 in *Moral Psychology: The Cognitive Science of Morality*. Vol. 2, edited by W. Sinnott-Armstrong. Cambridge, MA: MIT Press.

Livengood, Jonathan. 2013. "Actual Causation and Simple Voting Scenarios." *Nous* 47(2):316–45.

Menzies, Peter. 2004a. "Causal Models, Token Causation, and Processes." *Philosophy of Science* 71(5):820–32. doi: 10.1086/425057.

Menzies, Peter. 2004b. "Difference Making in Context." Pp. 341–67 in *Causation and Counterfactuals*, edited by J. Collins, N. Hall, and L. A. Paul. Oxford: Oxford University Press.

Menzies, Peter. 2007. "Causation in Context." in *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, edited by H. Price and R. Corry. Oxford University Press.

Menzies, Peter. 2017. "The Problem of Counterfactual Isomorphs." in *Making a Difference: Essays on the Philosophy of Causation*, edited by H. Beebee, C. Hitchcock, and H. Price. Oxford University Press.

Menzies, Peter, and Helen Beebee. 2019. "Counterfactual Theories of Causation" edited by E. Zalta. *Stanford Encyclopedia of Philosophy*.

Paul, L. A., and Ned Hall. 2013. *Causation: A User's Guide*. Oxford University Press.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Second edition., 3rd printing.. Cambridge, United Kingdom ; New York: Cambridge University Press.

Schaffer, Jonathan. 2004. "Causes Need Not Be Physically Connected to Their Effects: The Case for Negative Causation." Pp. 197–216 in *Contemporary Debates in Philosophy of Science*, *Contemporary Debates in Philosophy*, edited by C. Hitchcock. Malden MA: Blackwell Publishing.

Weslake, Brad. 2015. “A Partial Theory of Actual Causation.” *British Journal for the Philosophy of Science*.

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

Woodward, James. 2016. “The Problem of Variable Choice.” *Synthese* 193(4):1047–72. doi: 10.1007/s11229-015-0810-5.

Yablo, Stephen. 2002. “De Facto Dependence.” *The Journal of Philosophy* 99(3):130–48.

Zangwill, Nick. 2011. “Negative Properties.” *Noûs* 45(3):528–56.