Causal Models and Causal Relativism

Jenn McDonald

Abstract A promising development in the philosophy of causation analyzes actual causation using structural equation models, i.e., “causal models”. This paper carefully considers what it means for an interpreted model to be accurate of its target situation. These considerations show, first, that our existing understanding of accuracy is inadequate. Further, and more controversially, they show that any causal model analysis is committed to a kind of *relativism* – a view whereby causation is a three-part relation holding between a cause, an effect, and something else. In particular, insofar as a causal model analysis construes causation mind-and-language independently, it must treat causation as relative to a specification of background possibilities – i.e., a ‘modal profile.’ Or, so I argue.

§1 Introduction

Cutting-edge work in the philosophy of causation uses the framework of causal models – structural equation models (SEMs) and directed acyclic graphs (DAGs) – to analyze actual causation.[1] An interesting feature of this approach is that it leads with the formalism,

---

For discussion and feedback that greatly improved this work, I am indebted to Justin Clarke-Doane, Stephen Neale, David Papineau, Graham Priest, Jonathan Schaffer, and Michael Strevens. Thanks also to audiences at the Society for the Metaphysics of Science workshop (Fall 2021) and the Philosophy of Science Biennial Conference (Pittsburgh, Fall 2022).

[1] For seminal discussion of these models, see Spirtes, Glymour, and Scheines ([1993] 2000), and Pearl ([2000] 2009).

allowing cooperative interlocutors to be of very different metaphysical persuasions. Two parties can agree on the matter of how to read actual causal relations off of a model, but disagree over in what those relations bottom out (in counterfactuals, for example, or in type-level causal dependencies), or whether they can be genuinely reduced at all. At the same time, cooperative parties can disagree over what they take to be the causal relata – events, facts, property instances, etc. This raises the question whether the use of this formal framework demands any particular, metaphysical commitments. Is there any claim to which a causal model analysis of actual causation must be committed? I argue that there is: any such analysis must be a *relativist* about causation, at least insofar as causation is realistically construed.

Generally, relativism treats causation as holding between a cause, an effect, and some third relatum. One thing causes another only relative to this third relatum, different specifications of which may be given by different accounts. For example, one might take *c* to cause *e* relative to < c^*, e^* >, where c^* and e^* are non-empty sets of alternatives for *c* and *e*, respectively (Schaffer 2005; 2012; Hitchcock 2011).[2] In this paper, I argue that a causal model analysis is committed to a version of relativism whereby *c* causes *e* relative to what I call a 'modal profile,' which indicates whether and how the various actual factors in the situation could have gone differently.

---

[2] Note that views of this kind are called *contrastivist*, any version of which will count as a relativist view under this formulation. For an alternative version that posits only causal contrasts, see (Hitchcock 1996b; 1995; 1996a; 1993).

This may not seem so surprising. A model represents a particular factor – such as an event or property instance – as one value of a multi-valued variable, with the remaining values representing a range of alternatives. Isn't this just a transparent commitment to a kind of relativism? In fact, not at all, as I explain in §2.

The paper is laid out as follows. In §2, I review a causal model analysis of actual causation. This includes a brief overview of structural equation models, an enumeration of the components of a causal model analysis, and a first attempt at precisifying what makes a model "accurate" of its target situation. However, in §3 I show why this first attempt is inadequate – a model on an interpretation can still be made accurate or inaccurate of the same situation. I argue that this is due to an as yet unrecognized element in how causal models represent – models represent their target situations only relative to a modal profile. An amendment to what makes for accuracy is required. However, the needed amendment directly leads to a problem for extant causal model analyses of actual causation, as illustrated in §4. Two responses are considered in §5, the best of which I ultimately defend – adopting the view whereby actual causation holds relative to a modal profile. Call this view "Causal Relativism." Only Causal Relativism permits a realistically construed analysis that comports with causal intuition. I conclude by suggesting how it offers a practical, unified methodology for guiding causal inquiry in various domains.

§2 Analyzing Actual Causation: A Review

§2.1 Causal Models and Their Interpretations

A structural equation model (SEM) is an ordered triple, $<\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{L}}>$, comprised of a signature, an assignment, and a linkage.[3] The *signature*, $\boldsymbol{\mathcal{S}}$, is a set of variables, $\boldsymbol{U} \cup \boldsymbol{V}$, each of which is mapped to a range of possible values, $\boldsymbol{R}(X \in \boldsymbol{U} \cup \boldsymbol{V})$. Each variable is classified as either *exogenous* ($\boldsymbol{U}$) or *endogenous* ($\boldsymbol{V}$) which indicates, roughly, that it represents an independent or dependent condition, respectively. Each variable represents a factor alongside a range of alternatives or contrasts. Factors are the causal relata – events, property instantiations, facts, etc. The SEM framework allows any choice here. Its neutrality on this question allows it to bypass traditional debates over the nature of causal relata. For perspicuity, though, I will focus on property instantiations. Consider, then:

Forest Fire On a hike through the forest, Kenny discards his lit match onto dry kindling. The kindling ignites and the fire spreads.

We can use a binary variable to represent Kenny dropping a lit match alongside the alternative of him dropping a dead match. Doing the same for the kindling being dry or wet, and with a fire starting or not produces three binary variables on the following interpretation.

$\mathcal{I}(\mathcal{M}_1)_{FF}$: *X(match)* = $\begin{cases} 1\ if\ lit \\ 0\ if\ dead \end{cases}$ *Y(kindling)* = $\begin{cases} 1\ if\ dry \\ 0\ if\ wet \end{cases}$

[3] This formalism follows Halpern (2000) and Blanchard and Schaffer (2017).

$$Z\ (fire) = \begin{cases} 1\ if\ starts \\ 0\ if\ doesn't\ start \end{cases}$$

In this example, ***R*** maps every variable to {*0*, *1*}. Whether a variable is exogenous or endogenous in a model will be fully explained in terms of the linkage. For now, we know that *X* and *Y* represent independent conditions, but *Z* is dependent. So, ***U*** = {*X, Y*}, ***V*** = {*Z*}.

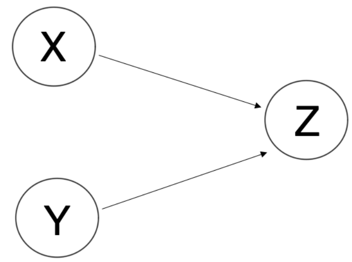
The next component of a SEM is the assignment, $\boldsymbol{\mathcal{A}}$, which assigns each exogenous variable, *X* ∈ ***U***, to one of its values, x_1 ∈ ***R***(*X*). Intuitively, the assignment represents the actual initial conditions of the situation. Here, the assignment should be *X* = *1* and *Y* = *1*, representing Kenny's dropping a lit match and the kindling being dry.

Finally, the linkage, $\boldsymbol{\mathcal{L}}$, is a set of asymmetric functional equations defined over the variables in $\boldsymbol{\mathcal{S}}$. The linkage indicates the dependence of the endogenous on the exogenous variables. It allows for representation of what actually happens as well as what would have happened had the alternatives occurred instead of what actually occurs. In **Fire**, the occurrence of fire depends on the match being lit and the kindling being dry. Had the match been dead or had the kindling been wet, the fire would not have started. This is captured by the equation, *Z := min(X, Y)*, which says that *Z* depends for its value on the values of *X* and *Y*, according to the function *min(X, Y)*. Combining this signature, assignment, and linkage, we get $\mathcal{M}_1$ which, interpreted using $\mathcal{I}(\mathcal{M}_1)_{FF}$, represents **Fire**:

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

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

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



$\mathcal{M}_1$

Equations are *asymmetric* in that they stipulate what value the left-hand variable (aka. the "child" variable) takes for any combination of values of the right-hand variables (aka. the "parent" variables), when the parent variables are set by *intervention*. An intervention is a surgical operation forcing only the specified variable to take one of its values, eliminating the variable's dependence on other variables, if any, and otherwise leaving the model as is.[4] More precisely, an intervention, $I_{X=x_i}$, on a variable, X, produces a sub-model identical to the original except that the X-equation is replaced with '$X = x_i$'. This renders X independent of its parent variables, but otherwise preserves the dependency structure of the model. Attention is restricted to *recursive* models with a unique equation for each endogenous variable. This means that the equations can be ordered such that once a variable appears on the right-hand side of an equation it will not again appear on the left-hand side, and that each equation can be labelled by its left-hand variable. For example, EQ3 is the Z-equation of $\mathcal{M}_1$.

Each equation represents some kind of dependence of the representatum of an endogenous variable on those of the other variables in the model. But the relevant dependence can be understood in different ways. Many take the equations to represent complex counterfactual

[4] This follows Pearl ([2000] 2009), see also (Briggs 2012). For a different formalization see (Woodward 2003).

dependencies. Researchers in this camp generally continue the tradition of reducing causal to counterfactual dependence (Hitchcock 2007a; 2009; Hall 2007; Hitchcock 2001), with some notable exceptions (Woodward 2003). Others take the equations to represent type-level causal dependencies, treated as primitive (Cartwright 2016; Hiddleston 2005; Pearl [2000] 2009; Gallow 2021) or reduced further into certain patterns of probabilistic dependencies (Papineau 2022). Most researchers of this stripe go on to give a counterfactual semantics in terms of these models, thereby reducing counterfactual dependence to type-level causal dependence (Briggs 2012; Hiddleston 2005; Pearl [2000] 2009). Elsewhere, equations are taken to represent sui generis structural determination relations (Gallow 2016; 2021), which are then used to give a semantics of causal counterfactuals. I maintain neutrality on this matter by putting what an interpreted model says in terms of the counterfactuals *entailed* by it – leaving open whether these counterfactuals are entailed because they are represented directly or they supervene on the represented structure. As I argue, the same problem arises regardless.[5] EQ3 can therefore be taken to entail the following counterfactuals:

$X = 1$ and $Y = 1$ (lit match and dry kindling) □→ $Z = 1$ (fire)

$X = 0$ (dead match) □→ $Z = 0$ (no fire)

[5] In fact, somewhat different counterfactuals will be entailed by a given interpreted model depending on what counterfactual semantics one supposes. An interventionist semantics (aka. a causal model semantics) diverges from a similarity semantics (say) in various ways – for example, in excluding counterfactuals with disjunctive antecedents (Pearl [2000] 2009; Halpern 2013; Hiddleston 2005), in assignment of truth-values to counterfactuals (Woodward 2003), or in what counts as a valid inference (Briggs 2012). (Although see (Vandenburgh forthcoming) for a possible way towards bringing the results of these two semantics back together.) This paper trades in examples selected so as to avoid this complication.

$Y = 0$ (wet kindling) $\square\!\!\rightarrow$ $Z = 0$ (no fire)

§2.2 Analyzing Actual Causation

From here, a SEM analysis of actual causation is constructed in three parts. The first specifies a "recipe" with which relations of actual causation can be read off a given model, such as what needs to hold of $< \mathcal{M}_1, \mathcal{I}(\mathcal{M}_1)_{FF} >$ in order for it to say that Kenny's dropping the lit match *caused* the fire. The recipe characterizes the structure of the causal relation in terms of what must hold of a model. For example, we might specify an actual causation relation as holding between the value of one variable and the value of a second whenever an intervention on the first leads to a change in the second. Formally:

Actual Causation *(AC)* $X = x$ is a cause of $Z = z$ relative to $\mathcal{M}_i$ iff in $\mathcal{M}_i$, (i) $X = x$ and $Z = z$, (ii) $X = x$ $\square\!\!\rightarrow$ $Z = z$, and (iii) $X = x'$ $\square\!\!\rightarrow$ $Z = z'$, where $x \neq x'$ and $z \neq z'$.

AC is, in fact, the core of all extant recipes. However, it cannot on its own handle *redundant causation* – cases presenting a back-up cause or an additional, independent cause. Amendment is therefore needed. However, further complexity only complicates the dialectic, leaving the substance of my argument unaltered.[6] So, I stick to ***AC*** as the basic recipe.

[6] Generally, more complex recipes incorporate a distinction between "on-path" and "off-path" variables, specifying that off-path variables should be held fixed somehow when checking for an effect of a putative cause variable on an effect variable. See, among others, (Hitchcock 2007a; Halpern and Pearl 2005; Weslake 2015; Halpern 2016a). A further, controversial complexity incorporates a normative parameter – which will generally both introduce additional conditions on how a model should be interpreted as well as revise the recipe to reflect

Notice that ***AC***, or any recipe, delivers only model-relative claims of actual causation. What we want, though, are claims of actual causation simpliciter. To get this, analyses quantify over the set of all appropriate, or "apt", models – on some to-be-determined notion of apt. The next part of any analysis, therefore, indicates a quantifier. I assume the existential quantifier.[7] The use of quantifier is precisely why a causal model analysis need not be committed to relativism simply due to its use of variables. While alternatives are utilized in a model's representation of a situation, whether or not *c* causes *e* merely depends on the existence of some set of alternatives such that the recipe is satisfied. It can then be said that *c* causes *e simpliciter*, not that *c* causes *e* relative to the qualifying set(s) of alternatives.

Naturally, this raises the question of what qualifies an interpreted model as *apt* for representing a given situation. The third and final part of an analysis, then, provides just such an account. I'll turn to this shortly. Taken together, a causal model analysis says that some property instantiation, *c*, causes another, *e*, just in case the right recipe is satisfied by an apt model of *c* and *e*.

the relevance of whether variables are set to their default or deviant settings. See, for example, (Hall 2007). Occasionally, the framework itself will be updated with a component responsible for tracking the default/deviant settings of variables (Gallow 2021).

[7] This choice is standard, with Hall's (2007) choice of the universal quantifier a notable exception. But the choice is merely conventional. Any choice of quantifier is in principle an option. It simply determines what work needs doing by one's theory of aptness. For example, given an existential quantifier, aptness needs to dispel model-interpretation pairs that deliver incorrect "true" verdicts. Given a universal one, it instead needs to dispel incorrect "false" verdicts.

§2.3 Defining Accuracy

While much ink has been spilled over the recipe,[8] less discussion has been had regarding aptness.[9] No complete and precise account of aptness can be found in the literature. It may even be a mistake to expect one, since aptness may be "more a matter of art than science" (Hitchcock 2007a, 503). Regardless, it is widely recognized that apt models are, at least, accurate ones.[10] However, what accuracy means precisely is left open. To remedy this, I begin with a definition of accuracy that systematizes and explicates what's been discussed in the literature. This stems from the basic idea that a model on an interpretation is accurate of some situation just in case it implies only true propositions about that situation. As a first pass, this will hold just in case the interpretation is permissible (on a to-be-defined notion of permissible), the values assigned to the exogenous variables represent actual property instances, and the dependencies or relations represented by the equations are *real* – in the simple sense that they really do hold of the situation. Formally,

[8] See, especially, (Weslake 2015; Halpern and Pearl 2005; Hall 2007; Hitchcock 2007a; Woodward 2003; Gallow 2021; Beckers and Vennekens 2018; Halpern 2016a).

[9] Of course, some progress has been made. For incomplete discussions of aptness, see (Halpern 2016b; 2016a; Hiddleston 2005; Hitchcock 2001; 2007a; 2012a; Menzies 2017; Gallow 2021; McDonald forthcominga); and for the most thorough work focused on aptness, see (Halpern and Hitchcock 2010; Woodward 2016; Hall 2007; Halpern 2016b; Blanchard and Schaffer 2017; Gallow 2016; McDonald 2024).

[10] Endorsement of some version of accuracy can be found in (Hitchcock 2001; Hall 2007; Blanchard and Schaffer 2017; Gallow 2016; 2021; Woodward 2003; Baumgartner 2013a; Paul and Hall 2013; Pearl and MacKenzie 2018).

Accuracy A causal model, $\mathcal{M}_i$, is accurate of a given situation, $\mathbb{S}$, on an interpretation $\mathcal{I}(\mathcal{M}_i)$, just in case

(1) $\mathcal{I}(\mathcal{M}_i)$ is a permissible interpretation of $\mathcal{M}_i$ for representing $\mathbb{S}$;

(2) The content entailed by the assignment, $\boldsymbol{\mathcal{A}}_{\mathcal{M}_i}$, on $\mathcal{I}(\mathcal{M}_i)$ is the case in $\mathbb{S}$;

(3) The dependencies represented by $\boldsymbol{\mathcal{L}}_{\mathcal{M}_i}$ on $\mathcal{I}(\mathcal{M}_i)$ really hold of $\mathbb{S}$.

Before evaluating, I need to say a bit more about (3) and clarify (1). As written, (3) leaves open the nature of the relevant dependencies. To remain neutral, as discussed earlier, I speak of the counterfactuals entailed by an interpreted model. So, (3) can be translated into: (3)' The counterfactuals entailed by $\boldsymbol{\mathcal{L}}_{\mathcal{M}_i}$ on $\mathcal{I}(\mathcal{M}_i)$ are true of $\mathbb{S}$.

Next, what counts as a *permissible* interpretation? I take my proposal, which follows, to regiment what can be found in the literature on variable selection. First, define an interpretation as an assignment of content to the variables in the manner laid out above. This assignment is governed by three widely presupposed yet rarely explored principles of variable selection – what I call *exclusivity*, *exhaustivity*, and *distinctness*. *Exclusivity* requires that the values of a single variable represent mutually exclusive property instantiations.[11] This ensures that a variable takes *at most* one of its values. *Exhaustivity* requires that a

[11] For references to exclusivity, see (Pearl [2000] 2009, 3; Woodward 2003, 98; Hitchcock 2004, 145; 2007b, 76; 2007a, 502; Briggs 2012, 142; Blanchard and Schaffer 2017, 182)

variable's values capture the entire range of alternatives for whatever property instantiation the variable represents.[12] This ensures that a variable takes *at least* one of its values.

Finally, *distinctness* holds that things which are represented by different variables should be relevantly independent.[13] How to precisify this remains open. Distinctness seems required here for the same reason as in the traditional counterfactual account of causation – to separate the wheat of causation from the chaff of mere counterfactual dependence or correlation. Causal dependence can be so distinguished only when we uniquely consider dependencies holding between *distinct* entities.[14] This avoids spurious causal relations popping up as the result of counterfactual dependence or correlations holding between things that are *conceptually* related (such as an apple's being red and its being crimson), *mereologically* related (such as the left-hand side of the table being made of wood and the whole table being made of wood), or *logically* related (such as it being the case that ϕ and it being the case that $\psi \wedge \phi$). But the question of distinctness is subtle.[15] There's a sense in

---

[12] For references to exhaustivity, see (Pearl [2000] 2009, 3; Hitchcock 2001, 287; Woodward 2016, 1064; Blanchard and Schaffer 2017, 182; Briggs 2012, 142)

[13] For references to distinctness, see (Hitchcock 2004, 146; Blanchard and Schaffer 2017, 182; Briggs 2012, 142; Paul and Hall 2013, 59; Hitchcock 2007a, 502; Baumgartner 2013a, 88). Distinctness is related to Woodward's principle of *independent fixability* – the requirement that any variable in a model be such that it can be fixed at any of its values without non-causally forcing any other variable to take a particular value (Woodward 2008; 2015; 2016, 1063–64). See also (Yang 2013, 330). A model whose set of variables satisfies independent fixability also satisfies distinctness. But see (Zhong 2020) for a defense of a strictly weaker principle.

[14] (D. Lewis 1973a; 2000; Kim 1974; Spirtes 2009).

[15] Note, too, that different notions of 'distinct' may be relevant in different applications of structural equation models. For example, if the application is in the service of discovering and understanding some other

which my decision to raise my arm may be *distinct* from the neurological underpinnings of that decision – as a non-reductive physicalist sees it, at least. Arguably, though, this isn't the kind of distinctness relevant to analyzing causation – of defining when one thing causes another.[16] Due to its complexity, a complete account cannot be defended here. For now, I assume the traditional understanding of distinctness that requires that the factors represented by values from different variables do not stand in any metaphysical dependence relations – such as conceptual, mereological, or logical dependence.[17]

Thus, an interpretation is *permissible* just in case whatever it says is exclusive, exhaustive, and distinct really *is* exclusive, exhaustive, and distinct. More precisely, $\mathcal{I}(\mathcal{M}_i)$, of a model, $\mathcal{M}_i$, is *permissible* for representing a situation, $\mathbb{S}$, if and only if the content that $\mathcal{I}(\mathcal{M}_i)$ assigns to the signature, $\boldsymbol{\mathcal{S}}_{\mathcal{M}_i}$, satisfies exclusivity, exhaustivity, and distinctness relative to $\mathbb{S}$.[18]

§3 An Additional Parameter: Accuracy as Relative to a Modal Profile

dependence relation – such as constitutive dependence. For some discussion, see (Alex Gebharter 2017; Baumgartner and Casini 2023).

[16] This is especially relevant to debates surrounding the causal exclusion problem. See, for example, (Baumgartner 2009; Woodward 2015; Alexander Gebharter 2017; Baumgartner 2013b).

[17] This account is especially tentative due to a recent argument to the effect that causal relations can hold between entities that nevertheless stand in non-causal dependency relations (Blanchard 2023; Friend 2019).

[18] I leave my analysis of permissibility here, but there may be need for further conditions. One such condition, *Serious Possibilities*, makes an appearance later on. This requires that only serious possibilities be represented, for some to-be-defined notion of serious. See (Blanchard and Schaffer 2017, 182; Hitchcock 2001, 287; Weslake 2015, 24; Woodward 2016, 1064). Another condition, and one that I won't discuss further, is that the property instances represented be intrinsically characterized (Blanchard and Schaffer 2017, 182).

So far, this systematizes what is already in the literature. But the utility of bringing it together lies in our now being able to see clearly why it falls short. As it turns out, accuracy is not fixed by a model, an interpretation, and a situation. Whether an interpretation is permissible and whether the entailed counterfactuals are true – that is, whether the represented dependencies really do hold – are each relative to an *additional* parameter: a specification of a space of background possibilities – what I call a 'modal profile.'

§3.1 Whether an Interpretation is Permissible

Take first the relativity of permissible interpretations. Consider:

> **Two Lamps** There are two lamps connected to a single light switch. Lamp-1 is on when the switch is up, and lamp-2 is on when the switch is down.[19]

In **Two Lamps**, it seems there can only be one closed circuit at a time. There is only so much copper wire, controlled by the switch, which closes one circuit or the other but not both – it isn't long enough to do so, and can't be in two places at once. So, the electrical current can only flow through one lamp at a time. As a result, one lamp being on is *mutually exclusive* with the other being on. To represent these property instantiations in a model, then, it is permissible to map them to two values of the same variable. Distinctness takes this further. As controlled by the switch, the copper wire can only be in one of two places. It not being in

[19] This example can be found in (Pearl [2000] 2009, 324; Weslake 2015, sec. 3.1).

one ipso facto means it must be in the other. Thus, the two property instantiations are not distinct. They *must* be represented by the same variable. This representation of **Two Lamps** treats certain actual features of the situation as given and not subject to variation: namely, the mechanism of the switch and the length of copper wire. Relative to this framing of the situation, representing lamp-1's being on and lamp-2's being on with a single variable is permissible, while doing so with two variables is impermissible. Call such a framing a 'modal profile', for reasons I explain shortly.

Yet it is easy to imagine that the aforementioned features *are* subject to variation. That is, we could frame the very same situation differently by allowing for different possibilities. In some sense, there could have been more copper wire present. If there had, then it would have been possible for both lamps to be on at the same time. So, even if there isn't *in fact*, there *could have been* enough wire to go around. If we allow that the amount of copper wire could vary in this way – again, not that it in fact varies, but that it *could* - then lamp-1 being on no longer necessitates anything about lamp-2 being on. This is a different modal profile from before, since it permits greater variability in how the situation could have gone. But relative to this new modal profile, the two lamps being on are distinct. Thus, distinctness permits them to be represented as values of different variables in a model. Exclusivity takes this further. With the possibility of additional wire, lamp-1's being on and lamp-2's being on are not necessarily mutually exclusive. Thus, they must be represented by different variables (insofar as they are both represented). Relative to this second modal profile, then, representing lamp-1's being on and lamp-2's being on with a single variable is *impermissible*, while doing so with two variables is *permissible*. But this is a different (in fact, the opposite) prescription than before. What counts as permissible, according to exclusivity and distinctness, is therefore relative to a modal profile.

Exhaustivity is also relative in this way. Consider:

> **Alice in the Factory (AF)** Alice the pigeon is trained to peck at and only at scarlet things. Alice lives in the yard of a paint chip factory that only produces scarlet and cyan chips. Alice sees a scarlet chip in the yard and pecks.[20]

Notice that **AF** is such that the paint chip's colors are constrained by the fact that the factory produces only two colors of chips: scarlet and cyan. So, relative to the modal profile constrained by any chip in question having been produced in the factory, the scarlet chip could only have otherwise been cyan. A binary variable that takes one value for scarlet and the other for cyan thus satisfies exhaustivity and is therefore *permissible*. But relative to what is physically possible, the paint chip could have been any color. The binary variable, {*scarlet*, *cyan*}, fails to satisfy exhaustivity relative to a modal profile constrained only by physical possibility, and is therefore *impermissible*. So, what counts as permissible, according to exhaustivity, is also relative to a modal profile.

§3.2 True Counterfactuals

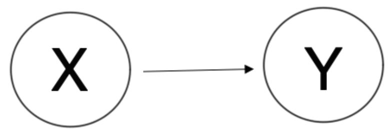
[20] This case is adapted from Shoemaker (2003), which is adapted from Yablo (1993).

Next, consider the matter of the entailed counterfactuals being true. Say we model **AF** with $\mathcal{M}_2$, on the interpretation $\mathcal{I}(\mathcal{M}_2)_{AF}$:

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

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

$\boldsymbol{\mathcal{L}}$ = (EQ1) $Y := X$



$\mathcal{M}_2$

$\mathcal{I}(\mathcal{M}_2)_{AF}$: $X\ (chip) := \begin{cases} 1\ if\ red \\ 0\ if\ not\ red \end{cases}$ $Y\ (Alice) := \begin{cases} 1\ if\ pecks \\ 0\ if\ doesn't\ peck \end{cases}$

Is $\mathcal{M}_2$ accurate of **AF** on $\mathcal{I}(\mathcal{M}_2)_{AF}$? Well, (1) is satisfied. The chip's being red and its being not red are exclusive and exhaustive alternatives, Alice pecking and not pecking are exclusive and exhaustive, and the chip's being red or not is distinct from Alice's pecking or not. Further, (2) is satisfied. The assignment sets *X* to 1, which represents the chip being red, which it is in **AF**. Finally, is (3)' satisfied? Here are the counterfactuals entailed by $\mathcal{M}_2$ on $\mathcal{I}(\mathcal{M}_2)_{AF}$:

(i) If the chip were red, then Alice would peck.

(ii) If the chip were not red, then Alice would not peck.

First, (ii) is straightforwardly true. The description of **AF** stipulates that Alice only pecks at scarlet things. Since there can be no non-red scarlet things, this means she will not peck at any non-red things. Is (i) true? Surprisingly, it depends. If we hold fixed the way this factory operates, then the only way a chip could be red in this factory yard is if it were scarlet. And if it were scarlet, then Alice would peck. So, when we allow what's possible to be constrained by contingent background facts, (i) comes out true. Therefore, $\mathcal{M}_2$ is accurate of **AF** on $\mathcal{I}(\mathcal{M}_2)_{AF}$ *relative to the modal profile constrained by how the factory actually operates*.

But it is not accurate tout court. If we allow that the paint chip could have been any physically possible color, (i) is false. If any alternative color is possible, then the chip could have been a non-scarlet shade of red. In this case, Alice would not have pecked. Thus, (i) is false. This means that $\mathcal{M}_2$ is *not* accurate of **AF** on $\mathcal{I}(\mathcal{M}_2)_{AF}$ *relative to the modal profile constrained only by physical possibility*.

§3.3 Modal Profiles

The upshot of this demonstration is that a model on an interpretation is accurate (or not) of a situation only *relative* to a modal profile.[21] A modal profile of a situation identifies the situation in part in terms of its modal features. The idea is that the same set of actual factors, comprising an actual situation, can be typed or individuated in different ways according to whether and how each factor could have occurred differently. Though not essential to the notion, a modal profile can be glossed in terms of possible worlds. A modal profile of a set of actual factors, or 'situation', is a set of worlds each member of which instantiates some version of that situation (or counterpart of that situation), where a "version" of the target

[21] This notion seems to be akin to Touborg's notion of 'possibility horizon' (Touborg 2018; Gunnemyr and Touborg 2023). There is also a kinship here with Gallow's (2016) discussion of relativity to "a region of modal space." However, this provides a different reason to consider the accuracy of a model as relative. Gallow's need for relativity derives from his requirement of "modularity," which I take to be either effectively replaced by the principles of variable selection I've outlined in the main text or else too demanding (Cartwright 2001; 2002).

situation is some variation on how this same situation could have taken place.[22] Crucially, a modal profile need not be the universal set of all such worlds.[23] Such a set tells, then, a partial story about how the target situation could have gone. We generate partial stories by holding fixed some set of facts about that situation, while allowing others to vary. The key observation is that holding fixed certain features of the actual situation will dictate how others can be allowed to vary.[24] If the background fact of there being only one length of copper wire is held fixed, it eliminates the possibility of both lamps being on at the same time.[25] Holding different features fixed will eliminate different possibilities. This means that, for any situation, there is a range of non-identical partial stories about how that situation could have gone. The indispensable role these partial stories play in evaluating counterfactuals has long been appreciated. For example, whether a counterfactual is true given a possible worlds semantics is a function not of what holds in every possible world in which the antecedent is true, but only in a specified subset of such worlds (D. Lewis 1973b; Stalnaker 1968).[26] I am now observing that they also play a crucial role in how a causal model represents, regardless of whether the underlying metaphysic is counterfactual.

[22] This raises the question of when an alternative set of factors or course of events counts as a version of some target situation. Note that this is a special case of the more general question about transworld identity (Mackie and Jago 2022). Without going into detail, I assume a fairly permissive answer to this question. But a precise answer is likely called for. Surprisingly, this means that causal models require engaging with traditional metaphysical questions more than some causal modelists had hoped.

[23] That it cannot be is argued for in §5.

[24] Why hold anything fixed? It turns out that, in order to represent certain situations, a causal model *must* hold something fixed. I discuss and demonstrate this in §5.

[25] Assuming, of course, that the general set-up of the lamps is also held fixed, that the laws of electromagnetism are held fixed, etc.

[26] *Which* subset being the key question (Veltman 2005; K. S. Lewis 2016; D. Lewis 1973c; Kratzer 2012; 1981; 1989; D. Lewis 1979).

§3.4 Updating Accuracy

The inadequacy of our earlier definition of accuracy has therefore been demonstrated. The same model under the same interpretation can be applied to a situation alongside one of two (or more) different modal profiles, with the interpretation permissible relative to one but impermissible relative to another, and/or the entailed counterfactuals being true relative to one but false relative to another. The modal profile is therefore, oddly enough, an *additional element of how causal models represent*. And yet, their relevance has gone largely unrecognized in the literature. Despite the widespread endorsement of accuracy as a necessary condition on aptness, accuracy's sensitivity to a modal profile has been overlooked.[27]

To amend this, I propose that specification of a modal profile be included as a further component of an interpretation of a causal model. A modal profile can be specified by explicitly identifying a subset of worlds or by enumerating the features of the situation supposed fixed. This updates the first and third accuracy conditions: $\mathcal{I}(\mathcal{M}_i)$ is a permissible interpretation of $\mathcal{M}_i$ for representing $\mathbb{S}$ just in case the content entailed by the signature, $\boldsymbol{\mathcal{S}}_{\mathcal{M}_i}$, given $\mathcal{I}(\mathcal{M}_i)$ satisfies exclusivity, exhaustivity, and distinctness *relative to the modal profile specified by* $\mathcal{I}(\mathcal{M}_i)$; and the counterfactuals entailed by $\boldsymbol{\mathcal{L}}_{\mathcal{M}_i}$ on $\mathcal{I}(\mathcal{M}_i)$ are true in $\mathbb{S}$ *relative to the modal profile specified by* $\mathcal{I}(\mathcal{M}_i)$.

[27] Arguably, the closest approximation to an acknowledgement of relativity to a modal profile can be found in discussions of *Serious Possibilities*. (See fn.14. and §6.1)

While I take this to be the most straightforward amendment, this is of course not the only amendment capable of incorporating the information inherent in a modal profile. For example, if what counts as relata is appropriately restricted, then information about the modal profile could be built directly into the original assignment of content to variables.[28] In order for an interpretation to count as permissible, on this view, each factor assigned to a given value of a variable must have its own modal profile built in. This may seem to more directly undermine the famous neutrality with respect to relata previously touted by causal model analyses (see §2.1). But the effect is, in fact, the same. Regardless of how the amendment is made, one upshot of this discussion is the importance of and consequent call to return to the metaphysic of causal relata – in particular, to its modal nature. Crucially, though, the call expands the scope beyond the cause and effect to the modal nature of features in the surrounding situation, as well.

§4 A Problem of Misleading Modal Profiles

Recognizing the need to incorporate a modal profile has widespread ramifications for causal inquiry. Of immediate interest is a problem it raises for analyses of actual causation. Since analyses quantify over all apt model-interpretation pairs, they quantify over any modal profile that figures in an apt model-interpretation pair. But very little has been said about what might constrain modal profiles, if anything. As it stands, all modal profiles are eligible

[28] Thanks to an anonymous referee from this journal for the alternative suggestion.

to figure in an apt model-interpretation pair. This is the problem: some deliver counterintuitive causal verdicts. I illustrate this with the choice of existential quantifier, though the problem arises in some form on any quantifier choice. There are three kinds of cases: overly general causes, irrelevant positive causes, and irrelevant omissive causes.

§4.1 Overly General Causes

First up is a case where an overly general property instantiation qualifies as a cause. Refer to the situation already provided when **Alice in the Factory** was modeled with <$\mathcal{M}_2$, $\mathcal{I}(\mathcal{M}_2)_{AF}$>. Since $\mathcal{I}(\mathcal{M}_2)_{AF}$ didn't include reference to a modal profile, specify that constrained by being in the factory yard. Call this $\mathcal{I}(\mathcal{M}_2)'_{AF}$. This is the modal profile relative to which the original model on the original interpretation was accurate. Thus, on the new interpretation that includes this, the model will be accurate of **AF**.

$\mathcal{M}_2$ says that *X = 1* is an actual cause of *Y = 1*, and $\mathcal{I}(\mathcal{M}_2)'_{AF}$ takes this to say that the chip's being *red* is an actual cause of Alice pecking. Since there is at least one apt model-interpretation pair that delivers this verdict, the chip's being red *just is* an actual cause of Alice pecking. It is an actual cause *simpliciter*. But this is counterintuitive. Being red is too general to be an actual cause. The chip's being red was causally efficacious only because it happened to be scarlet, due to its being in the factory yard. But it could have been red, in a natural sense of "could", without being scarlet. And then Alice would *not* have pecked. The verdict is at minimum highly misleading.[29]

§4.2 Irrelevant Positive and Omissive Causes

A modal profile can also qualify a prima facie irrelevant property instantiation as a cause: either an irrelevant *positive* property instantiation or an irrelevant *omissive* property instantiation.[30] The following can serve to illustrate both kinds of case.

> **The Prince and his Biscuits (PB)** The Queen of England has to be out for the day. She asks the Prince of Wales to water her plant in her absence. Prince agrees, but eats biscuits instead of watering the plant. The plant wilts.

[29] A similar argument shows how overly *specific* property instantiations qualify as causes, but it strikes me as less compelling. Briefly, imagine Sophie the pigeon is trained to peck at *red* things (not scarlet, like Alice). Use $\mathcal{M}_2$ to model this along the lines of $\mathcal{I}(\mathcal{M}_2)_{AF}$, but replace Alice with Sophie and re-interpret X so 1 represents the chip's being scarlet and 0 not scarlet. Add specification of a modal profile constrained by Sophie's being in the factory yard. This model-interpretation pair is accurate and, according to it, the chip's being scarlet caused Sophie to peck. Thus, the chip's being scarlet *just is* a cause of Sophie pecking. But this result is overly specific. In a natural enough sense of "could", the chip could have been a non-scarlet shade of red. If so, Sophie would still have pecked. Given this, it is false that had the chip not been scarlet, then Sophie would not have pecked. That being said, this result strikes me as less misleading than the others. But disagreement on this point simply makes for more ways modal profiles could mislead, strengthening the force of the problem.

[30] These problems come from (Sartorio 2010), where they are presented as two problems for counterfactual accounts of causation generally: the *'Problem of Unwanted Positive Causes'* and the '*Problem of Unwanted Negative Causes'*. **The Prince and his Biscuits** is adapted from an example Sartorio provides.

Suppose also that there is a locking mechanism on the greenhouse that unlocks the room from 12:00 to 12:15, and this coincides with the only time of day that biscuits are put out in the tearoom on the far side of the palace. It takes at least twenty minutes to get from the greenhouse to the tearoom, or back again.

Take first the problem of irrelevant positive causes. **PB** can be accurately modelled using $\mathcal{M}_2$ on the following interpretation:

$\mathcal{I}(\mathcal{M}_2)_{PB-P}$: *X (Prince of Wales) :=* $\begin{cases} 1\ if\ eats\ biscuits \\ 0\ if\ waters\ plant \end{cases}$ *Y (plant) :=* $\begin{cases} 1\ if\ wilts \\ 0\ if\ thrives \end{cases}$

Modal Profile: constrained by lock mechanism and palace layout

$\mathcal{I}(\mathcal{M}_2)_{PB-P}$ is permissible relative to the specified modal profile, the assignment says truly that Prince ate biscuits, and the counterfactuals entailed by $\boldsymbol{\mathcal{L}}_{\mathcal{M}_2}$ are true. So, <$\mathcal{M}_2$, $\mathcal{I}(\mathcal{M}_2)_{PB-P}$> is accurate. However, according to $\mathcal{M}_2$, *X = 1* actually causes *Y = 1*, and on $\mathcal{I}(\mathcal{M}_2)_{PB-P}$ this means that Prince's eating biscuits is an actual cause of the plant wilting. So, Prince's eating biscuits *just is* an actual cause simpliciter of the plant wilting. But this is counterintuitive! Prince's eating biscuits seems irrelevant. It is causally efficacious to the wilting of the plant only because it excludes his watering, due to the layout of the palace and locking mechanism. In an obvious sense, he *could* have watered the plant while eating biscuits. Had he, the plant would not have wilted. This result is misleading, at best.

Finally, to illustrate the problem of irrelevant *omissive* causes, suppose the biscuits give Prince a stomachache. This can be accurately modelled using $\mathcal{M}_2$, on the following interpretation:

$\mathcal{I}(\mathcal{M}_2)_{PB-O}$: *X (Prince of Wales) :=* $\begin{cases} 1\ if\ doesn't\ water\ plant \\ 0\ if\ waters\ plant \end{cases}$

Y (Prince's stomach) := $\begin{cases} 1\ if\ aches \\ 0\ if\ doesn't\ ache \end{cases}$

Modal Profile: constrained by lock mechanism and palace layout

However, according to $\mathcal{M}_2$, *X = 1* actually causes *Y = 1*, and on $\mathcal{I}(\mathcal{M}_2)_{PB-O}$ this means that Prince's not watering the plant is an actual cause of him developing a stomachache. So, Prince's not watering the plant *just is* an actual cause of his developing a stomachache. It is an actual cause simpliciter. This, too, seems counterintuitive. Prince's not watering the plant is prima facie irrelevant to his developing a stomachache. Prince instantiating the property of not watering the plant is causally efficacious only because it is coextensive with his eating biscuits, given the locking mechanism and the palace layout. In some sense, though, he could have watered the plant *and* eaten the biscuits. Had that been the case, he still would have developed a stomachache. Furthermore, there is a sense in which he could have failed to water the plant while also failing to eat the biscuits. Had that been the case, he would not have developed a stomachache. This result is highly misleading.[31]

---

[31] I should flag that the foregoing treatment of the misleading cases presupposes the notion of accuracy previously laid out and, in particular, the notion of distinctness as *metaphysically* distinct. A different construal of distinctness may call for a different treatment of these cases – if, for example, relata represented by distinct variables were permitted to be metaphysically intertwined. Thanks to an anonymous referee for this journal

§5 Relativity as Sleight of Hand?

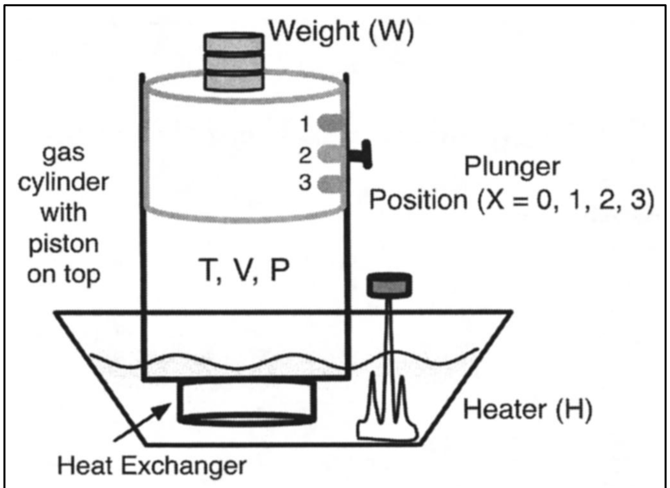
Before turning to responses, it will help to dispel a concern one might have with the argument so far. One might worry that interpreted models only appear to represent relative to a modal profile because the focus has been on simple models capable of representing only a local, limited portion of reality. A *universal* model-interpretation pair – a fully exhaustive interpreted model that represents the one true story about all the possible ways things could have been – would be accurate of any situation *tout court*. The relativity is therefore an illusion brought about by our needing to use simpler model-interpretation pairs due to various cognitive limitations. If so, there is no metaphysical motivation for incorporating such relativity into a theory of aptness.

There are two reasons to think this can't be right. First, a universal model-interpretation pair is not possible. A concise argument for this is given by Hausman, Stern, and Weinberger (2014) in the context of DAGs, but the point generalizes to SEMs. They show how altering the initial conditions of some physical systems changes the very structure of that system. Structural changes of this kind cannot be captured within a single model-interpretation pair. Consider their example of **The Elusive Cylinder**:

---

for emphasizing this point. Even so, I suspect the need for modal profile would still arise. While the implications of tweaking this parameter merits further investigation, space constraints unfortunately prohibit exploring this here.

[The device] consists of gas immersed in a water bath that is maintained at a constant temperature H. There is a [plunger] at the top of the cylinder that can be locked into one of three positions (*X = 1, 2,* or, *3*) or allowed to move up or down depending on the pressure of the gas (*X = 0*) and on the weight placed on top of the [plunger]. (Hausman, Stern, and Weinberger 2014, 1926)



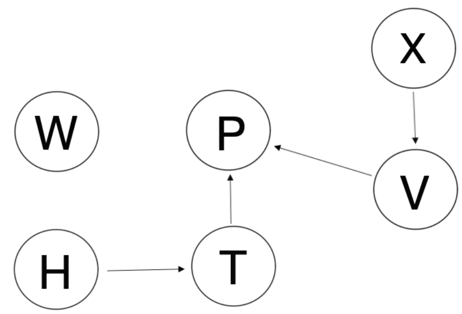
The Elusive Cylinder (Hausman et al., 2014, p. 1926)

The key feature is that the dependency relations that hold between the property instantiations in this system when the plunger is locked at some one of its possible positions *structurally* differ from when it is not locked. This means that an intervention that changes the plunger from locked to unlocked, or back again, will not simply propagate throughout a single model's system of equations, changing the values of variables in line with the equations. This is how an intervention normally goes. Instead, an intervention on the plunger makes for a change in the equations themselves, rearranging which variables determine which others. To illustrate, compare how a set of equations and corresponding graph would accurately represent, on the one hand, the situation assuming that the plunger is locked and, on the other, the situation assuming that the plunger floats freely. For both, the signature is basically the same: the set of exogenous variables, $\boldsymbol{U} = \{H, W, X\}$ and the set of endogenous variables, $\boldsymbol{V} = \{P, T, V\}$. The interpretation of these variables is indicated already in the figure, above. We can ignore both the particular values and the assignment.

$\boldsymbol{\mathcal{S}}$ = $\boldsymbol{U} = \{H, W, X\}$; $\boldsymbol{V} = \{P, T, V\}$; $\boldsymbol{R}$ = *n/a*

$\boldsymbol{\mathcal{A}}$ = *n/a*

$\boldsymbol{\mathcal{L}}_{X \neq 0}$ = (EQ1) $T := f_T(H)$
(EQ2) $P := f_P(T, V)$
(EQ3) $V := f_V(X)$

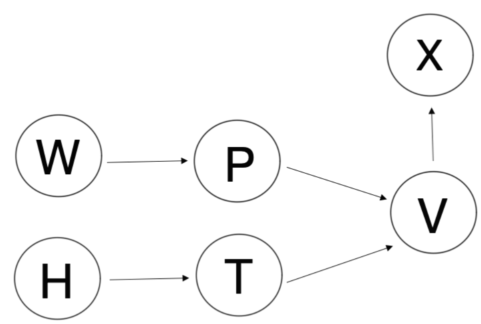


$\mathcal{M}_3$

$\boldsymbol{\mathcal{S}}$ = $\boldsymbol{U} = \{H, W, X\}$; $\boldsymbol{V} = \{P, T, V\}$; $\boldsymbol{R}$ = *n/a*

$\boldsymbol{\mathcal{A}}$ = *n/a*

$\boldsymbol{\mathcal{L}}_{X = 0}$ = (EQ1) $T := f_T(H)$
(EQ2) $P := f_P(W)$
(EQ3) $V := f_V(P, T)$
(EQ4) $X := f_V(V)$



$\mathcal{M}_4$

$\mathcal{I}(\mathcal{M}_{3\ \&\ 4})$: *W (Weight); H (Heater); X (Piston/Plunger); T (Temperature); P (Pressure); V (Volume)*

Crucially, the structures of the two models are different, as indicated by different equations and graphs. When the plunger is locked ($\mathcal{M}_3$), the weight plays no role, the locking position of the plunger fully determines the volume of the gas, which in turn affects the pressure. But when the plunger is not locked ($\mathcal{M}_4$), its location is fully a function of the volume of the gas, which in turn is determined by the pressure and temperature, the former of which is affected by the weight. So, the roles of various variables diverge, as does the direction of dependence. But this describes two distinct structures possibly manifest by the *same* physical system. Suppose the plunger is unlocked and focus on the model that represents this system accurately ($\mathcal{M}_4$). It cannot simultaneously represent the same system with a locked plunger. And yet, there is still the *possibility* that the plunger be locked. Thus, there exists a possibility for how this system could have gone that cannot be represented in the same model that represents its actual conditions. The general upshot is that some possible changes in features of the world change the relationships between other features. In model terms, this means that some interventions on variables change the form of the equations. But this cannot be captured within a single model in the SEM formalism. There can be no such thing as a universal model-interpretation pair.

But perhaps there is some other way we can meaningfully represent the one true story about all possible ways things could go. Assume there is for the sake of argument. The second problem with this line of reasoning is that using such a universal system of interpreted models would make it so that almost every property-instance is a direct cause of almost every other.[32] This is recognized by Spirtes, Glymour, and Scheines in their discussion of type-level causation:

> If our notion of causation between variables were strictly applied, *almost every natural variable would count as a [type-level] cause of almost every other natural*

[32] Thanks to Jonathan Schaffer for suggesting this response.

variable, for no matter how remote two variables, A and B, may be, there is usually *some* physically possible – even if very unlikely – arrangement of systems such that variation in some values of A produces variation in some values of B. (1993, p. 44, emphasis my own)

The point applies to actual causation, as well. Take almost any two property instantiations, F_1a and G_1b, however intuitively unrelated, where G_1b lies in F_1a's forward light cone. There will be an alternative *albeit perhaps highly unlikely* property, F_2, that a could have instantiated instead such that, had it been the case that F_2a, then G_2b would have been the case, where $G_1 \neq G_2$ and G_1 and G_2 are exclusive. For example, take a particular car being parked on May 21, 1932 in the suburbs of Seattle, WA and the COVID-19 pandemic beginning in December 2019 in Wuhan, China. Had that car instead instantiated the property of time-travelling through a wormhole to Wuhan and exploding in the right spot and right time to kill patient zero before the virus spread, then the pandemic would not have happened. While a wildly remote possibility, the instantiation of such a property is nevertheless consistent with our best physical theories. If *all* possibilities are to be included, then so should this. The result would be that this car being parked on this day in this spot is a direct actual cause of the COVID-19 pandemic. While the idea of its being an *indirect* cause of the pandemic can perhaps be wrestled with – by positing a long enough chain of causal intermediaries – the force of the implication lies in the verdict of its being a *direct* cause. Indeed, every property instantiation in the backwards light cone of the dawn of COVID-19 would be a direct actual cause of the pandemic. In general, a universal system of interpreted models would make it so that F_1a is a direct actual cause of *every* property instance in its forward light cone, and every property instance in F_1a's backwards light cone is one of its direct actual causes.[33]

§6 If You Can't Beat Them, Join Them

So, we have a problem of modal profiles delivering counterintuitive results. Two responses merit consideration. The first resists these results by supplementing aptness so as to rule out offending modal profiles. The second embraces them – taking the relativity at face value. The view that follows is one whereby actual causation itself holds relative to a modal profile, with causal claims interpreted as involving an implicit reference to a modal profile. Between them, only the second is compatible with a realist construal of causation.

§6.1 Supplementing Aptness – Can We Get Serious?

A natural response here is to supplement our theory of aptness specifically to rule out those modal profiles relative to which counterintuitive causal verdicts come out. This whittles down the domain of modal profiles over which the SEM recipe quantifies. The resulting view would comport with intuition, assuming its success. One would also hope it could preserve the *objectivity* of causation – in the sense that questions about causation have determinate answers. That is, by quantifying over all apt model-interpretation pairs, one would think

[33] The severest version of this problem follows from the assumption that *backwards* time travel is also possible. The result would be that every feature of reality is both a direct actual cause and a direct effect of every other feature. Whether backwards time travel is consistent with our best physical theories remains a point of controversy, however. So, I keep to the milder problem in the text.

such a view could deliver a determinate answer to the question of what causes what, and thereby preserve objectivity about causation.[34] Note that this may read as a peculiar use of 'objective.' Objectivity is, perhaps, more standardly run together with *realism* – i.e., the idea that questions about an area have mind- and language-independent answers. Aside from substantive reasons to treat them as separate,[35] by so doing I can be more precise about how the current response differs from the one to follow.

In fact, there already exists an aptness principle in the literature seemingly suited for this work. *Serious Possibilities* enjoins an interpreted model to represent only possibilities that are serious (Blanchard and Schaffer 2017, 182; Hitchcock 2001, 287; Woodward 2016, 1064). Invoking it, one could attempt to identify non-serious possibilities represented by the problematic model-interpretation pairs. Already, though, this is not quite right. Some earlier counterintuitive verdicts arise from an overly conservative admission of possibilities, rather than an overly liberal one. They go wrong not by including non-serious possibilities, but in omitting serious ones. Yet this suggests a simple fix: an interpreted model should represent *all and only* those possibilities that are serious. Call this *Serious Possibilities**. One can now argue that in representing **AF** as a situation in which paint chips can only be scarlet or cyan, $<\mathcal{M}_2, \mathcal{I}(\mathcal{M}_2)_{AF}>$ omits the serious possibility of a paint chip being some other color. Similarly, in representing **PB** as a situation in which the Prince can either water the plant or else eat

[34] For this way of defining 'objectivity,' and so distinguishing it from 'realism,' see (Clarke-Doane 2020, 27).

[35] Again, see (Clarke-Doane 2020). There is also the simple fact that 'objective' is notoriously ambiguous, and begs stipulation anyhow.

biscuits, both $< \mathcal{M}_2,\ \mathcal{I}(\mathcal{M}_2)_{PB-P} >$ and $< \mathcal{M}_2,\ \mathcal{I}(\mathcal{M}_2)_{PB-O} >$ omit the serious possibility of him doing both. Thus, these are all inapt. Problem solved.

I have left the notion of *serious* intentionally opaque, to highlight the promise of this response. However, delivery on this promise requires the notion be clarified. As I see it, this cannot be done satisfactorily. To begin, *Serious Possibilities** is too strong in requiring *all* serious possibilities be represented. There is a sense in which any *real* possibility – as in, any non-epistemic possibility – counts as serious. On this sense, though, nothing short of a fully exhaustive model could satisfy this principle. However, no such model is possible and even if an exhaustive representation were somehow engineered, it would deliver the verdict that everything is a direct cause of everything else. This was shown in §5. So, what counts as "serious" needs restriction.

The route forward suggested by the literature is treacherous, at least for a realist. Blanchard and Schaffer, for example, employ the idea of possibilities that "we are willing to take seriously" (2017, 197) to solve the problem of omissions. The gardener's failure to water the plants is on par, dependency-wise, with the Queen's failure to water them. Why, then, is it the gardener's failure (and not the Queen's) that causes them to die? The answer is that the Queen's watering the plants is not a serious possibility, and so a model that represents her watering them violates *Serious Possibilities(*)*. Fair enough. But *why* aren't we willing to take it seriously? Arguably, because it is highly *unlikely* that the Queen would water the plants, or because the Queen's watering the plant is highly *unusual* or *atypical*. So, it comes down to a measure of the *likelihood* or *normality* of a possibility that determines whether it counts as serious.

But we are not out of the woods. A likelihood measure or one of normality has its own problems. First, a means of delivering categorical measures is not forthcoming. There are various reasons for this, from the reference class problem (Hájek 2007), to the context-sensitivity of counterfactuals (K. S. Lewis 2016; K. Lewis 2018), to counterfactual skepticism (Emery 2017), to the sheer variety and incomparability of normative considerations (Blanchard and Schaffer 2017). So much for the objectivity of causation on this response.

Worse, however, is the plausible requirement that pragmatic considerations will need to figure into any measure of likelihood or normality whose causal verdicts successfully comport with intuition. Evidence from psychology and experimental philosophy suggest that causal judgment is systematically affected by normative considerations, including moral norms and norms of convention (Knobe and Fraser 2008; Hitchcock and Knobe 2009). A likelihood measure will need to take these into account in order to capture such judgments, but incorporating norms in this way gives up on realism about causation – not to mention provides for a psychologically implausible account of causal judgment (Blanchard and Schaffer 2017). Alternatively, one could insist on purely non-pragmatic considerations figuring into the measure of likelihood or normality, thereby preserving realism at the cost of comporting with a range of causal intuitions. Such a view would still need to respond to the charge of psychological implausibility.

In sum, an aptness principle like *Serious Possibilities** does not actually preserve objectivity as was hoped, and either commits us to anti-realism about causation or fails to fully capture our causal intuitions. This is not damning. Arguably, any metaphysical theory conflicts with some degree of intuition. Causal judgment and intuition is, after all, known to be biased and fallible.[36] In addition, some in the literature are already happy to relinquish realism about actual causation (Hall 2007; Halpern and Hitchcock 2010; Hitchcock and Knobe 2009). However, I think this is both too great a concession and too quick. There is another response that preserves realism while better capturing our intuitions, or so I will argue.

§6.2 Causal Relativism

The more promising response, to my mind, takes this relativity at face value, incorporating it by positing modal profile as a third relatum in the causal relation.[37] Consider carefully the counterintuitive verdicts previously laid out. It does seem wrong to say that the chip's being red is an actual cause *simpliciter* of Alice pecking. But does this mean it is *wrong* simpliciter? There is a sense in which the chip's being red is not an actual cause of Alice pecking. But there

---

[36] For example, work in psychology supports an *outcome-density* or *outcome-frequency bias* – an increase in our tendency to judge an event or action as causal when the effect occurs more frequently (Alloy and Abramson 1979; Musca et al. 2010; Allan and Jenkins 1980; 1983; Wasserman et al. 1996; Buehner, Cheng, and Clifford 2003) – as well as a *cue-density bias* – an increase that manifests when the putative cause occurs more frequently (Allan and Jenkins 1983; Wasserman et al. 1996; Vadillo et al. 2011; Blanco, Matute, and Vadillo 2013). Causal perception also seems to give rise to causal illusions, arguably due to its being informationally encapsulated (Beebee 2009; Scholl and Tremoulet 2000; Schlottmann 2000; Schlottmann and Shanks 1992).

[37] See also (Gunnemyr and Touborg 2023; Touborg 2018) for a view of this kind, argued to from somewhat different considerations.

is also a sense in which it *is*. It makes sense to say that the chip's being red is *not* an actual cause of Alice pecking given the possibility that the chip could have been red without being scarlet. But it also makes sense to say that the chip's being red *is* an actual cause of Alice pecking given the impossibility of any red chip in the factory yard failing to be scarlet. There is truth to both claims. It strikes me that the real problem with existentially quantifying over modal profiles is not that it mistakenly includes those that deliver counterintuitive results, but that it mistakenly omits a crucial part of what makes causal claims true – namely, the background possibilities relative to which causation holds.

Applying this to **The Prince and his Biscuits**, it is true that the Prince's eating biscuits is not an actual cause simpliciter of the plant wilting. There is no causation simpliciter! Instead, the Prince's eating biscuits is an actual cause of the plant wilting given the lock mechanism and layout of the palace – which makes it impossible for him to both eat biscuits and water the plant. Given the physical possibility of him doing both, though, his eating biscuits is not an actual cause of the plant wilting. Similarly, the Prince's not watering the plant is an actual cause of his stomachache given the lock mechanism and layout of the palace, but not given the physical possibility of him watering the plant *and* eating biscuits. That is all there is to it.

So, I propose that actual causation is relative in nature. On this view, the counterintuitive verdicts are explained by positing relativity to modal profile as an implicit parameter in causal claims. The causal claim "the prince's eating biscuits is an actual cause of the plant wilting" is underspecified. Fill it in with different modal profiles to produce different truth-values and correspondingly different causal intuitions.

Still, one might see some modal profiles as preferable. Aren't some simply more legitimate than others? Indeed! In fact, engagement with this question is independently triggered by the relativist view on the table. In seeking to make explicit what has otherwise been a hidden parameter in causal claims, questions naturally arise about which modal profiles are of interest and why. There is evidence, for example, that we have a preference in causal judgment for causal relations that are portable and robust, supporting accurate predictions and guiding successful behavior across many kinds of situations without requiring the careful tracking of background conditions (Lombrozo 2010; Quillien 2020; Hitchcock 2012b).[38] Relations of this sort will hold relative to modal profiles constrained only by those contingent facts which commonly hold in everyday environments. Causal claims relative to modal profiles constrained by peculiar contingent facts (such as the locking mechanism and layout of the palace) will be unreliable unless such facts are specially tracked, increasing cognitive load.

We just saw (in §6.1), however, the difficulty of explaining in mind-and-language independent terms what makes some modal profiles more "legitimate". It is a key advantage of the current view that this preference can be explained in the obvious way – as due to the pragmatic benefit incurred. And yet, this invocation of pragmatic considerations in no way threatens the mind-and-language independence of causation. Once we fix on a modal profile,

[38] See also (Woodward 2001; Hitchcock and Woodward 2003; Woodward and Hitchcock 2003) for arguments to the effect that portability (what they call *invariance*) plays a significant role in causal explanation.

it is in no sense *up to us* what causes what. Instead, what is up to us is *which of the many different causal structures we attend to*.

In order to preserve realism about causation, however, this view gives up on objectivity. Causation is no longer objective in the sense that there is no uniquely correct causal structure. Instead, there are many different structures, relative to each of which different actual causation relations may hold. There are no categorical facts as to what actually causes what. Determinate facts about what actually causes what are relative to a modal profile. Of course, the previous attempt to supplement aptness also gives up on objectivity. So, this is not a relative disadvantage. However, we might take this to be a mark against any view to the extent we believe that causation is a determinate matter full stop. That being said, it strikes me as no great loss since determinacy is recovered once the modal profile is given. In particular, in the event of tension between realism and objectivity, realism strikes me as the feature worth having.

The final advantage of this view, which I call "Causal Relativism", is that it provides a realism-friendly account of the effect of norms on causal judgment. In response to the evidence mentioned above, many argue we should incorporate norms into our metaphysics.[39] For those of us who want to preserve realism, though, Causal Relativism provides an alternative treatment of the evidence. Normative considerations can be translated in terms of modal

[39] See (Gallow 2021; Hall 2007; Halpern and Hitchcock 2010; Halpern 2016b; Halpern and Hitchcock 2015; Menzies 2017).

profiles, and the effect of norms on causal judgment can be straightforwardly explained as a pragmatic preference for some modal profiles over others.[40]

§7 Conclusion

I've shown how the accuracy of a model is relative to an interpretation – one which includes specification of a modal profile – and a situation. However, quantifying over all modal profiles delivers counterintuitive results. Attempts to dispel these results by strengthening our account of aptness threatens to undermine realism about causation. The most viable response, as I have argued, is instead to treat actual causation as relative to a modal profile, which makes for a kind of relativism I term "Causal Relativism". This response provides a practical methodology for explaining causal judgment without needing to give up on realism.

While discussion has been kept short, other advantages and ramifications of Causal Relativism merit further discussion. For want of space, I name three. First, it arguably has all the benefits of any contrastivist view of causation, plausibly figuring as a demanding "total contrasts" version of contrastivism (Schaffer 2005; 2012). In addition, the modal profile plays a clear role in filling in the content of negations. For **AF**, for example, relative to the first modal profile – the one constrained by how the factory operates – 'not-red' refers to cyan. Relative to the second one – the one constrained by physical possibility – 'not-red' refers to all non-red colors. This will likely dictate answers to questions surrounding causation by omission. Finally, it can serve as an independently motivated metaphysical

[40] See (McDonald forthcomingb) for expansion on this application of Causal Relativism.

basis for the linguistic framework recently proposed by Touborg (2022) as a way of defending the principle of strong proportionality.[41] In light of these, further exploration of the view is clearly warranted.

§8 References

Allan, L. G., and H. M. Jenkins. 1980. "The Judgment of Contingency and the Nature of the Response Alternatives." *Canadian Journal of Experimental Psychology* 34 (1): 1–11.

———. 1983. "The Effect of Representations of Binary Variables on Judgment of Influence." *Learning and Motivation* 14:381–405.

Alloy, L. B., and L. Y. Abramson. 1979. "Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?" *Journal of Experimental Psychology: General* 108 (4): 441–85.

Baumgartner, Michael. 2009. "Interventionist Causal Exclusion and Non-reductive Physicalism." *International Studies in the Philosophy of Science* 23 (2): 161–78. https://doi.org/10.1080/02698590903006909.

———. 2013a. "A Regularity Theoretic Approach to Actual Causation." *Erkenntnis* 78:85–109.

———. 2013b. "Rendering Interventionism and Non-Reductive Physicalism Compatible." *Dialectica* 67 (1): 1–27.

[41] See (McDonald 2022) for a defense along similar lines.

Baumgartner, Michael, and Lorenzo Casini. 2023. "The PC Algorithm and the Inference to Constitution." *British Journal for the Philosophy of Science* 74 (2): 405–29.

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

Beebee, Helen. 2009. "Causation and Observation." In *The Oxford Handbook of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Peter Menzies. Oxford University Press.

Blanchard, Thomas. 2023. "The Causal Efficacy of Composites: A Dilemma for Interventionism." *Philosophical Studies* 180:2685–2706.

Blanchard, Thomas, and Jonathan Schaffer. 2017. "Cause Without Default." In *Making a Difference: Essays on the Philosophy of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Huw Price, 175–214. Oxford University Press. https://doi.org/10.1093/oso/9780198746911.003.0010.

Blanco, F., H. Matute, and M. A. Vadillo. 2013. "Interactive Effects of the Probability of the Cue and the Probability of the Outcome on the Overestimation of Null Contingency." *Learning and Behavior* 41 (4): 333–40.

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

Buehner, M. J., P. W. Cheng, and D. Clifford. 2003. "From Covariation to Causation: A Test of the Assumption of Causal Power." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29 (6): 1119–40.

Cartwright, Nancy. 2001. “Modularity: It Can - and Generally Does - Fail.” In *Stochastic Causality*, edited by Maria Carla Galavotti, Patrick Suppes, and Domenico Costantini, 65–84. CSLI Publications.

———. 2002. “Against Modularity, the Causal Markov Condition, and Any Link Between the Two: Comments on Hausman and Woodward.” *British Journal for the Philosophy of Science* 53:411–53.

———. 2016. “Single Case Causes: What Is Evidence and Why.” In *Philosophy of Science in Practice: Nancy Cartwright and the Nature of Scientific Reasoning*, 11–24. Springer.

Clarke-Doane, Justin. 2020. *Morality and Mathematics*. Oxford University Press.

Emery, Nina. 2017. “The Metaphysical Consequences of Counterfactual Skepticism.” *Philosophy and Phenomenological Research* 94 (2): 399–432.

Friend, Toby. 2019. “Can Parts Cause Their Wholes.” *Synthese* 196:5061–82.

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

———. 2021. “A Model-Invariant Theory of Causation.” *Philosophical Review*.

Gebharter, Alex. 2017. “Uncovering Constitutive Relevance Relations in Mechanisms.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 174 (11): 2645–66.

Gebharter, Alexander. 2017. “Causal Exclusion and Causal Bayes Nets.” *Philosophy and Phenomenological Research* 95:353–75.

Gunnemyr, Mattias, and Caroline Touborg. 2023. “You Just Didn’t Care Enough.” *Journal of Ethics and Social Philosophy* 24 (1): 1–35.

Hájek, Alan. 2007. "The Reference Class Problem Is Your Problem Too." *Synthese* 156:563–85.

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

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

———. 2013. "From Causal Models to Counterfactual Structures." *The Review of Symbolic Logic* 6 (2): 305–22.

———. 2016a. *Actual Causality*. MIT Press. http://www.jstor.org.ezproxy.gc.cuny.edu/stable/j.ctt1f5g5p9.

———. 2016b. "Appropriate Causal Models and the Stability of Causation." *The Review of Symbolic Logic* 9 (01): 76–102. https://doi.org/10.1017/S1755020315000246.

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

———. 2015. "Graded Causation and Defaults." *The British Journal for the Philosophy of Science* 66 (2): 413–57.

Halpern, Joseph, 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. https://doi.org/10.1093/bjps/axi147.

Hausman, Daniel M., Reuben Stern, and Naftali Weinberger. 2014. "Systems without a Graphical Causal Representation." *Synthese* 191 (8): 1925–30.

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

Hitchcock, Christopher. 1993. “A Generalized Probabilistic Theory of Causal Relevance.” *Synthese* 97 (3): 335–64.

———. 1995. “The Mishap at Reichenbach Fall: Singular vs. General Causation.” *Philosophical Studies* 78 (3): 257–91. https://doi.org/10.1007/BF00990114.

———. 1996a. “Farewell to Binary Causation.” *Canadian Journal of Philosophy* 26:267–82.

———. 1996b. “The Role of Contrast in Causal and Explanatory Claims.” *Synthese* 107 (3): 395–419.

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

———. 2004. “Routes, Processes, and Chance-Lowering Causes.” In *Cause and Chance: Causation in an Indeterministic World*, edited by Phil Dowe and Paul Noordhof. Routledge.

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

———. 2007b. “What’s Wrong with Neuron Diagrams?” In *Causation and Explanation*, edited by J. K. Campbell, M. O’Rourke, and H. S. Silverstein, 4–69. MIT Press.

———. 2009. “Causal Modelling.” In *The Oxford Handbook of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Peter Menzies, 1:299–314. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199279739.003.0015.

———. 2011. “Trumping and Contrastive Causation.” *Synthese* 181 (2): 227–49.

———. 2012a. “Events and Times: A Case Study in Means-Ends Metaphysics.” *Philosophical Studies* 160 (1): 79–96.

———. 2012b. “Portable Causal Dependence: A Tale of Consilience.” *Philosophy of Science* 79:942–51.

Hitchcock, Christopher, and Joshua Knobe. 2009. “Cause and Norm.” *Journal of Philosophy* 106 (11): 587–612.

Hitchcock, Christopher, and James Woodward. 2003. “Explanatory Generalizations, Part II: Plumbing Explanatory Depth.” *Noûs* 37 (2): 181–99.

Kim, Jaegwon. 1974. “Noncausal Connections.” *Noûs* 8 (1): 41–52. https://doi.org/10.2307/2214644.

Knobe, Joshua, and Ben Fraser. 2008. “Causal Judgment and Moral Judgment: Two Experiments.” In *Moral Psychology: The Cognitive Science of Morality*, edited by Walter Sinnott-Armstrong, 2:441–47. Cambridge, MA: MIT Press.

Kratzer, Angelika. 1981. “Partition and Revision: The Semantics of Counterfactuals.” *Journal of Philosophical Logic* 10 (2): 201–16.

———. 1989. “An Investigation of the Lumps of Thought.” *Linguistics and Philosophy* 12 (5): 607–53.

———. 2012. *Modals and Conditionals: New and Revised Perspectives*. Oxford University Press.

Lewis, David. 1973a. “Causation.” *The Journal of Philosophy* 70 (17): 556. https://doi.org/10.2307/2025310.

———. 1973b. *Counterfactuals*. Cambridge, MA: Harvard University Press.

———. 1973c. “Counterfactuals and Comparative Possibility.” *Journal of Philosophical Logic* 2 (4): 418–46.

———. 1979. “Counterfactual Dependence and Time’s Arrow.” *Nous* 13 (4): 455–76.

———. 2000. “Causation as Influence.” *The Journal of Philosophy* 97 (April):182–97. https://doi.org/10.2307/2678389.

Lewis, Karen. 2018. “Counterfactual Discourse in Context.” *Noûs* 52 (3): 481–507.

Lewis, Karen S. 2016. “Elusive Counterfactuals.” *Noûs* 50 (2): 286–313.

Lombrozo, Tania. 2010. “Causal-Explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions.” *Cognitive Psychology* 61:303–32.

Mackie, Penelope, and Mark Jago. 2022. “Transworld Identity.” Edited by Edward N. Zalta and Uri Nodelman. *The Stanford Encyclopedia of Philosophy* Fall 2022 Edition. URL = <https://plato.stanford.edu/archives/fall2022/entries/identity-transworld/>.

McDonald, Jennifer. 2022. *Actual Causation: Apt Causal Models and Causal Relativism*. The Graduate Center, CUNY. http://academicworks.cuny.edu/gc_etds/4828.

———. 2024. “Causal Models and Metaphysics - Part 2: Interpreting Causal Models.” *Philosophy Compass* 19 (7).

———. forthcominga. “Essential Structure for Causal Models.” *Australasian Journal of Philosophy*.

———. forthcomingb. “Norms in Actual Causation.” *Erkenntnis*.

Menzies, Peter. 2017. “The Problem of Counterfactual Isomorphs.” In *Making a Difference: Essays on the Philosophy of Causation*, edited by Helen Beebee, Christopher Hitchcock, and Huw Price. Oxford University Press.

Musca, S. C., M. A. Vadillo, F. Blanco, and H. Matute. 2010. “The Role of Cue Information in the Outcome-Density Effect: Evidence from Neural Network Simulations and a Causal Learning Experiment.” *Connection Science* 22 (2): 177–92.

Papineau, David. 2022. “The Statistical Nature of Causation.” *The Monist* 105:247–75.

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

Pearl, Judea. (2000) 2009. *Causality: Models, Reasoning, and Inference*. Second edition., 3rd printing.. Cambridge, United Kingdom ; New York: Cambridge University Press.

Pearl, Judea, and D. MacKenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.

Quillien, Tadeg. 2020. "When Do We Think That X Caused Y?" *Cognition* 205.

Sartorio, Carolina. 2010. "The Prince of Wales Problem for Counterfactual Theories of Causation." In *New Waves in Metaphysics*, edited by A. Hazzlett, 259–76. New York: Palgrave Macmillan.

Schaffer, Jonathan. 2005. "Contrastive Causation." *Philosophical Review* 114 (3): 327–58. https://doi.org/10.1215/00318108-114-3-327.

———. 2012. "Causal Contextualism." In *Contrastivism in Philosophy: New Perspectives*, edited by Martijn Blaauw. Routledge.

Schlottmann, Anne. 2000. "Is Perception of Causality Modular?" *Trends in Cognitive Science* 4:441–42.

Schlottmann, Anne, and David Shanks. 1992. "Evidence for a Distinction between Judged and Perceived Causality." *Quarterly Jounal of Experimental Psychology* 44:321–42.

Scholl, B. J., and P. D. Tremoulet. 2000. "Perceptual Causality and Animacy." *Trends in Cognitive Science* 4:299–309.

Spirtes, Peter. 2009. "Variable Definition and Causal Inference." *Proceedings of the 13th International Congress of Logic, Methodology, and Philosophy of Science*, 514–53.

Spirtes, Peter, Clark N. Glymour, and Richard Scheines. (1993) 2000. *Causation, Prediction, and Search*. Lecture Notes in Statistics (Springer-Verlag) ; v. 81. New York: Springer-Verlag.

Stalnaker, Robert. 1968. "A Theory of Conditionals." *American Philosophical Quarterly*, 98–112.

Touborg, Caroline. 2018. *The Dual Nature of Causation*. PhD Thesis, University of St Andrews.

———. 2022. "Relativizing Proportionality to a Domain of Events." *Synthese* 200 (2): 1–20.

Vadillo, M. A., S. C. Musca, F. Blanco, and H. Matute. 2011. "Contrasting Cue-Density Effects in Causal and Prediction Judgments." *Psychonomic Bulletin and Review* 18:110–15.

Vandenburgh, Jonathan. forthcoming. "Backtracking Through Interventions: An Exogenous Intervention Model for Counterfactual Semantics." *Mind & Language*.

Veltman, Frank. 2005. "Making Counterfactual Assumptions." *Journal of Semantics* 22 (2): 159–80.

Wasserman, E. A., S. F. Kao, L. J. Van Hamme, M. Katagiri, and M. E. Young. 1996. "Causation and Association." In *Causal Learning*, 34:207–64. The Psychology of Learning and Motivation. San Diego, CA: Academic Press.

Weslake, Brad. 2015. "A Partial Theory of Actual Causation." *Manuscript*.

Woodward, James. 2001. "Law and Explanation in Biology: Invariance Is the Kind of Stability That Matters." *Philosophy of Science* 68:1–20.

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

———. 2015. “Interventionism and Causal Exclusion.” *Philosophy and Phenomenological Research* 91 (2): 303–47. https://doi.org/10.1111/phpr.12095.

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

Woodward, James, and Christopher Hitchcock. 2003. “Explanatory Generalizations, Part I: A Counterfactual Account.” *Noûs* 37 (1): 1–24.

Yang, Eric. 2013. “Eliminativism, Interventionism and the Overdetermination Argument.” *Philosophical Studies* 164 (2): 321–40. https://doi.org/10.1007/s11098-012-9856-0.