Apt Causal Models and the Relativity of Actual Causation

Jenn McDonald

Abstract Recent work promises to analyze actual causation using causal models. Any such analysis must include how a model should map onto the world. A natural thought is that a model must at least be accurate – saying only true things. However, I argue that this is overly simple. I demonstrate how accuracy is not had tout court, but only relative to a space of possibilities. This discovery raises a problem for extant causal model theories and, indeed, any theory of actual causation in terms of counterfactual or type-level causal dependence. I conclude with a view that resolves this problem.

§1 Introduction

A recent development in the philosophy of causation uses the framework of causal models, such as structural equation models, to analyze actual causation. There are two components to such an analysis. The first analyzes the causal relation in terms of a model or class of models.[1] The second provides an account of what qualifies a model or class of models as *apt* such that the analysis can be taken to make commitments about the real world. A complete and precise account of aptness cannot be found in the literature and doesn't seem

[1] See (Halpern and Pearl 2005; Hitchcock 2001; Weslake 2015; Woodward 2003; Gallow 2021; Beckers and Vennekens 2018; Hall 2007; J. Y. Halpern 2016a).

forthcoming.[2] But it is universally agreed that an apt model must at least be *accurate*, where accuracy is understood along the following lines: a model on an interpretation is accurate of a situation when it says only true things about that situation. I will show, however, that this is overly simplistic. I first propose a standard method of interpretation and define accuracy assuming this method. I then demonstrate how 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 space of possibilities. This raises a problem for extant causal model analyses of actual causation, which I illustrate. But the problem is general. It affects any analysis of actual causation in terms of whatever kind of underlying dependence these models might be taken to represent. Such theories include any that seeks to reduce actual causation to counterfactual or type-level causal dependence, regardless of whether it itself invokes the causal model framework. I conclude with a general view of actual causation that solves this problem.

§2 Causal Models

Consider the following.

[2] For discussion of aptness, see (Blanchard and Schaffer 2017; Hall 2007; J. Halpern and Hitchcock 2010; J. Y. Halpern 2016b; 2016a; Hiddleston 2005; Hitchcock 2001; 2007a; 2012; Menzies 2017; Woodward 2016).

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

We can represent this with a *structural equation model* (SEM), $< \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{A}}, \ \boldsymbol{\mathcal{L}} >$. First, factors such as property instantiations or events are represented by variables. Each actual factor is represented alongside a range of alternatives for how that factor could have otherwise occurred by the values of a given variable. The set of variables constitutes the *signature*, $\boldsymbol{\mathcal{S}}$ = {***U***, ***V***, ***R***}.[3] It consists in a set of *exogenous* variables, ***U***, which are independent variables, a set of *endogenous* variables, ***V***, which are dependent ones, and a relation, ***R***, that maps each variable onto a range of values. Let's represent Kenny dropping a lit match versus a dead one as the two values of a binary variable. The same applies for the kindling being dry or wet and with a fire starting or not. This produces a model with three binary variables and their interpretation:

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

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

In this example, ***U*** = {*X, Y*}, ***V*** = {*Z*}, and ***R*** maps each variable to {*0*, *1*}.

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

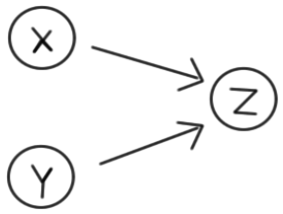
Next, the assignment, $\boldsymbol{\mathcal{A}}$, maps to each exogenous variable, $X \in \boldsymbol{U}$, one of its values, $x_1 \in \boldsymbol{R}(X)$. Intuitively, the assignment represents the initial conditions. Here, the assignment is $X = 1$ and $Y = 1$, representing that Kenny dropped a lit match and that the kindling is dry.

Finally, the set of asymmetric functional equations defined over the variables from the signature is the *linkage*, $\boldsymbol{\mathcal{L}}$. The linkage represents the dependence of the endogenous on the exogenous variables. It captures what actually happens as well as what would have happened had the alternatives occurred instead. In **Fire**, had the match been dead or had the kindling been wet, the fire would not have started. This is captured by the equation, $Z := min(X, Y)$.

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

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

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



$\mathcal{M}_1$

The equations 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.[4] 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 '$X=x_i$'. Such an operation renders X independent of its parent variables, but otherwise preserves the dependency structure of the model.

The literature divides on whether the dependencies represented by equations are counterfactual or causal.[5] I remain neutral by saying that counterfactuals are *entailed* by a model – leaving open whether they are entailed because they are directly represented or because they follow from the type-level causal structure that is represented. As we'll see, the same problem arises regardless. EQ3 can therefore be taken to entail the following counterfactuals:

$X = 1$ and $Y = 1$ (lit match and dry kindling) $\Box\!\!\rightarrow Z = 1$ (fire)

$X = 0$ (dead match) $\Box\!\!\rightarrow Z = 0$ (no fire)

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

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

[5] The former includes (Hitchcock 2007a; 2009; Woodward 2003), while the latter includes (Cartwright 2016; Hiddleston 2005; Pearl 2000). Proponents of the former generally though not always continue the tradition of seeking to reduce causal dependence to counterfactual dependence, while those who take the latter option treat type-level causal dependence as primitive.

Analyses of actual causation in terms of these models specify a recipe with which relations of actual causation can be read off a given model. Consider:

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

AC is the core of any extant recipe.[6] While amendment is still required to accommodate redundant causation, I set this complication aside.[7]

§2 Accuracy

Notice that ***AC*** produces only model-relative claims of actual causation. To arrive at claims of actual causation simpliciter, theories generally *existentially* quantify over the set of appropriate, or *apt*, models.[8] While aptness seems to be "more a matter of art than science (Hitchcock 2007a, 503)," it is nevertheless universally agreed that apt models must at least be accurate ones.[9] To evaluate this proposal, though, accuracy must first be defined. I propose that a model on an interpretation is accurate of some situation just in case it says

[6] See fn. 1.

[7] Redundant causation involves causal dependence without counterfactual dependence, such as preemption and overdetermination.

[8] This is the general choice, although (Hall 2007) selects a universal quantifier.

[9] See fn. 2.

only true things about that situation. As a first pass, this holds just in case the interpretation is permissible, the values assigned to the exogenous variables represent actual factors, and the entailed counterfactuals are true. More formally,

> **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 counterfactuals entailed by $\boldsymbol{\mathcal{L}}_{\mathcal{M}_i}$ on $\mathcal{I}(\mathcal{M}_i)$ are true in $\mathbb{S}$.

Allow me to explain 'permissible interpretation' before demonstrating why this account of accuracy is inadequate. I define an interpretation as an assignment of content to the variables. This assignment is governed by three widely presupposed yet rarely explored principles of variable selection – what I will call exclusivity, exhaustivity, and distinctness. *Exclusivity* requires that the values of a single variable represent mutually exclusive things.[10] This ensures that a variable takes *at most* one of its values. *Exhaustivity* requires that a variable's values capture the entire range of alternatives for whatever thing the

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

variable represents.[11] This ensures that a variable takes *at least* one of its values. With these two principles, variables become partitions on modal space.

Distinctness then holds that things which are represented by different variables should be relevantly distinct.[12] How to precisify this remains open. Distinctness is needed here for the same reason as in a traditional counterfactual account of causation – to separate the wheat of causation from the chaff of mere counterfactual dependence. Causation as counterfactual dependence only works when qualified as holding between *distinct* entities.[13] This avoids spurious causal relations popping up as the result of counterfactual dependence that holds between things that are *conceptually* related (such as an apple's being red depending on it being crimson), *mereologically* related (such as the left-hand side of the table being made of wood depending on the whole table being made of wood), or *logically* related (such as it being the case that ϕ depending on it being the case both that $\psi \rightarrow \phi$ and that ϕ). Roughly, then, distinctness requires that no two values from any two variables should stand in any conceptual, mereological, or logical dependency relations with each other.

[11] See (Pearl 2000, 3; Hitchcock 2001, 287; Woodward 2016, 1064; Blanchard and Schaffer 2017, 182; Briggs 2012, 142)

[12] See (Hitchcock 2004, 146; Blanchard and Schaffer 2017, 182; Briggs 2012, 142; Paul and Hall 2013, 59; Hitchcock 2007a, 502)

[13] (Lewis 1973; 2000; Kim 1974)

Given these principles, 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, $\mathbb{S}$, iff the content assigned to $\boldsymbol{S}_{\mathcal{M}_i}$ by $\mathcal{I}(\mathcal{M}_i)$ satisfies exclusivity, exhaustivity, and distinctness relative to $\mathbb{S}$.

§3 Accuracy as Relative

It turns out, though, that accuracy is not a determinate function of a model, an assignment of content to its variables, and a situation. Whether an interpretation is permissible and whether the entailed counterfactuals are true are each relative to an additional parameter – a specification of a *space of background possibilities*.

Take first the relativity of the permissibility of an interpretation. Consider:

> **Lamps** Two lamps connect to a single light switch. Lamp-1 is on when the switch is up, and lamp-2 is on when the switch is down.[14]

In **Lamps**, there can only be one closed circuit at a time – meaning the electrical current can only flow through one lamp at a time. Relative to the space of possibilities entailed by this, one circuit's being closed is *mutually exclusive* with the other being closed, and the current flowing through lamp-1 is *mutually exclusive* with it flowing through lamp-2. Thus, the two

[14] This comes from (Pearl 2000, 324; Weslake 2015, sec. 3.1).

events of lamp-1 being on and lamp-2 being on can be represented by the same variable and still satisfy exclusivity. Distinctness takes this further. Since they are not distinct, they *must* be represented by the same variable. There is only one bit of copper wire, controlled by the switch, which can close one circuit or the other but not at the same time – it cannot be in two places at once. Thus, it is not possible for the current of electricity to flow through lamp-1 while flowing through lamp-2, nor vice versa. Lamp-1 being on forces lamp-2 to be off, and lamp-2 being on forces lamp-1 to be off. Relative to this first space of possibilities, 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.

But it is easy to imagine relaxing this constraint, which would then make it possible that both lamps be on at the same time. It would merely require introducing another bit of copper wire to close the second circuit – all other conditions could be held the same. Relative to this second space of possibilities, the current flowing through one circuit and so illuminating one lamp does not necessitate anything about whether current flows through the other circuit and so illuminates the second lamp. Since the two lamps being on are *distinct* in this way, distinctness permits them to be represented as values of different variables. Exclusivity takes this further. Since lamp-1's being on and lamp-2's being on are *not mutually exclusive*, they must be represented by different variables (insofar as they are both represented). Relative to this second space of possibilities, 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*.

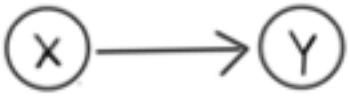
Exhaustivity is also relative. Consider:

> **Alice** Alice the pigeon pecks only at scarlet things. She lives in the yard of a paint chip factory that only produces scarlet and cyan chips. Alice sees a scarlet chip and pecks.[15]

Relative to the space of possibilities constrained by being in the factory yard, 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* relative to this first space of possibilities. But relative to what is broadly physically possible, the paint chip could have been any color. The binary variable, {*scarlet*, *cyan*}, fails to satisfy exhaustivity relative to this second space of possibilities, and is therefore *impermissible*.

Next, consider the relativity of the accuracy of the equations. Say we model **Alice** 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) $Y := X$

$\mathcal{M}_2$

[15] Adapted from Shoemaker (2003), which is adapted from Yablo (1993).

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

Is $\mathcal{M}_2$ accurate of **Alice** 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 **Alice**. 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 true. Intervening to set the chip to not red would result in Alice not pecking. Is (i) true? Surprisingly, it depends. First notice that **Alice** is such that the paint chip's colors are constrained by something in the background. The factory produces only two colors of chips: scarlet and cyan. 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 **Alice** on $\mathcal{I}(\mathcal{M}_2)_{AF}$ *on the space of possibilities constrained by how the factory actually operates*.

But it isn't accurate tout court. If we allow that the paint chip could have been any physically possible color, (i) is false. Many permissible interventions on the situation would

set the chip to a non-scarlet shade of red, in which case Alice would not have pecked, rendering (i) false. Thus, $\mathcal{M}_2$ is *not* accurate of **Alice** on $\mathcal{I}(\mathcal{M}_2)_{AF}$ *on the space of possibilities constrained by physical possibility*.

§4 Modal Profiles

So, accuracy of a model on an interpretation is relative to a *space of possibilities*. I call this a *modal profile*. This is a slight revision on the term "modal profile" as it's standardly used. "Modal profile" generally refers to the full range of possibilities of a thing. The following quote is illustrative: "A modal profile…captures all the possible combinations of properties the object might instantiate in different possible worlds. (Schroeter 2019, n. 2)" In this sense, the "modal profile" of a situation is the full story of how things in that situation could have been or gone otherwise. But notice that holding certain features fixed will rule out incompatible pieces of the story. When we hold fixed the background fact of there being only one bit of copper wire, this rules out the possibility of both lamps being on at the same time. Holding different features fixed will rule out different pieces of the story. This means that for any situation there is a whole range of *partial* stories about how that situation could have gone, each member of which results from the holding fixed of some set of facts about that situation. It has long been appreciated that these partial stories play a crucial role in the evaluation of counterfactuals. I am here demonstrating that they play a crucial role in how a causal model represents, as well. Due to their significance, it will be convenient to permit the term *modal profile* to refer to a *partial* range of possibilities for

how that situation could have gone. In my sense, then, there is not a single modal profile of a thing but a family of them.

While I won't fully account for the nature of modal profiles here, it may be helpful to briefly put the notion in terms of possible worlds. The *universal modal profile* of a situation is the set of worlds each member of which instantiates a version of the situation in question – which is some variation on how this situation could have taken place. Holding fixed certain features of the situation amounts to taking a proper subset of these worlds – only those which instantiate versions of this situation that are consistent with the features supposed fixed. "Modal profile" refers to a range of possible property-instantiations of that thing in *a certain specified subset* of possible worlds. A modal profile can be specified by explicitly selecting a subset of worlds or by enumerating the features of a situation supposed fixed.

So, the same model under the same interpretation can be applied to a situation alongside one of two (or more) different modal profiles. Relative to one it is accurate, while it is inaccurate relative to the other. The modal profile is therefore, strangely enough, an as yet *unrecognized, additional element* of how causal models represent. I propose it be incorporated by including a specification of modal profile as part of the interpretation.

§5 Problem of Counterintuitive Verdicts

This relativity to modal profile has widespread ramifications. Most interesting is a problem it raises for extant theories of actual causation in terms of these models. The first thing to

note is that by existentially quantifying over all apt model-interpretation pairs, these theories existentially quantify over any modal profile that figures in an apt model-interpretation pair. Since relativity to modal profiles has not been previously recognized, nothing has been said about them with regards to aptness. As it stands, then, any modal profiles is eligible to figure in an apt model-interpretation pair. However, this produces a problem. Some modal profiles deliver counterintuitive causal verdicts. There are four kinds of such verdicts: overly general causes, overly specific, irrelevant positive causes, and irrelevant omissive. I'll illustrate the first and third. It should be clear how to generate the others.

As an illustration of overly general causes, refer back to **Alice** modeled with $\mathcal{M}_2$ on $\mathcal{I}(\mathcal{M}_2)_{AF}$. Since $\mathcal{I}(\mathcal{M}_2)_{AF}$ didn't yet include a specification of modal profile, say we add the specification of the modal profile 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, the model will be accurate of **Alice**.

Now notice that $\mathcal{M}_2$ says that *X = 1* is an actual cause of *Y = 1*, and $\mathcal{I}(\mathcal{M}_2)'_{AF}$ interprets this to say that the chip's being *red* is an actual cause of Alice pecking. And 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*.

However, this result seems counterintuitive. Red is too general to be a cause. The chip's being red was causally efficacious only because it happened to be scarlet, due to being in the factory yard. But there's a sense in which it could have been red without being scarlet and, had that been the case, then Alice would *not* have pecked. This result is at minimum highly misleading.

To illustrate irrelevant positive causes, consider the following situation.

> **Prince's Biscuits (PB)** The Queen of England has to be out. She asks the Prince of Wales to water her plant. The Prince agrees, but eats biscuits instead. The plant dies.[16]

Suppose the following is also the case. The greenhouse unlocks only from 12pm to 12:20, which coincides with the only time of day biscuits are put out in the tearoom on the far side of the palace. It would take the fastest runner 20 minutes to get from the greenhouse to the tearoom, or back again. This can also be accurately modelled using $\mathcal{M}_2$, using the following interpretation:

$\mathcal{I}(\mathcal{M}_2)_{PB}$: *X (Prince)* := $\begin{cases} 1 \; if \; eats \; biscuits \\ 0 \; if \; waters \; plant \end{cases}$ *Y (plant)* := $\begin{cases} 1 \; if \; dies \\ 0 \; if \; survives \end{cases}$

Modal Profile: constrained by the lock mechanism, biscuit schedule, and palace layout

[16] This example comes from (Sartorio 2010).

$\mathcal{M}_2$ on $\mathcal{I}(\mathcal{M}_2)_{PB}$ is accurate for representing **PB**. The assignment says truly that the Prince ate biscuits, and, relative to the specified modal profile, both $\mathcal{I}(\mathcal{M}_2)_{PB}$ is permissible and the counterfactuals entailed by $\mathcal{L}_{\mathcal{M}_2}$ are true. But according to $\mathcal{M}_2$, *X = 1* actually causes *Y = 1*. $\mathcal{I}(\mathcal{M}_2)_{PB}$ interprets this to mean that the Prince's eating biscuits is an actual cause of the plant dying. So, since there is at least one apt model-interpretation pair that delivers this verdict, the Prince's eating biscuits *just is* an actual cause of the plant dying. It is an actual cause *simpliciter*.

However, this also seems counterintuitive. The Prince's eating biscuits is irrelevant to the dying of the plant. His eating biscuits was only causally efficacious because it happened to exclude his watering the plant, due to the layout of the palace and locking mechanism. But there's a sense in which he could have watered the plant while eating biscuits and, had that been the case, the plant would not have died. This result is also highly misleading.

§6 Actual Causation as Relative to Modal Profile

There are several possible responses to this problem. But I want to consider one in particular. What happens if we take seriously this relativity to modal profile, treating the metaphysical relation of actual causation as itself holding relative to modal profile?

This response can be motivated by considering more carefully the counterintuitive verdicts previously laid out. It seems wrong to say that the chip's being red is an actual cause *simpliciter* of Alice pecking. But this does not necessarily mean that it is *wrong* simpliciter

that the chip's being red is an actual cause of Alice pecking. Indeed, there is a sense in which the chip's being red is not an actual cause of Alice pecking. But there 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 metaphysical 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 contingent fact that any red chip in the factory yard is a scarlet chip*. It strikes me that the real problem with existentially quantifying over all modal profiles is that it omits a crucial part of the story – namely, what background possibilities are in place.

Applying this to **Prince's Biscuits**, it seems wrong to say that the Prince's eating biscuits is an actual cause *simpliciter* of the plant dying. But it makes sense to say that the Prince's eating biscuits *is* an actual cause of the plant dying *given the fact constrained by the lock mechanism and layout of the palace*, and that the Prince's eating biscuits *is not* an actual cause of the plant dying *given the fact that it is metaphysically possible for him to both eat biscuits and water the plant.*

I propose that actual causation itself holds relative to modal profile. On this view, the counterintuitive feel to the verdicts delivered by our theory can be explained by positing relativity to modal profile as a hidden parameter in our causal claims and by making this parameter explicit. The causal claims, "the chip's being red is an actual cause of Alice pecking" and "the prince's eating biscuits is an actual cause of the plant dying," are underspecified. Filling each in with different modal profiles will produce different causal intuitions.

Indeed, this view of actual causation opens up what strikes me as a potentially fruitful line of inquiry. 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. Upon examination of everyday causal claims, for example, we will likely discover a preference for causal relations that are highly portable and robust, supporting accurate predictions and guiding successful behavior without requiring the careful tracking of background conditions. Relations of this sort will hold relative to those modal profiles that are constrained only by those contingent facts which commonly hold in everyday environments. Causal claims relative to modal profiles constrained by highly peculiar contingent facts will be unreliable unless such peculiar contingent facts are tracked, increasing cognitive load.

It is an advantage of the proposed view that this preference can be explained in the obvious way – 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, it is in no sense *up to us* what causes what. Instead, what is up to us is which of the many different possible real underlying causal structures we attend to.

While this view preserves *realism* about causation, it does so by giving up on the *objectivity* of causation. Causation is no longer objective in the sense that there is no uniquely correct

causal structure.[17] There are many different correct structures, relative to each of which different actual causation relations may hold. The view is therefore a kind of *causal relativism*. There is no categorical fact as to what actually causes what. Determinate facts about what actually causes what are instead relative to a modal profile. This is a disadvantage insofar as we find compelling the claim that causation is determinate full stop. But it strikes me as no great loss given determinacy is recovered once the modal profile is filled in.

§7 Conclusion

A model's accuracy is relative to an interpretation – one which includes specification of modal profile – and a situation. However, existentially quantifying over all modal profiles produces counterintuitive causal verdicts. In response, I propose we take actual causation itself to be relative to modal profile, resulting in causal relativism. I have argued that this view best captures our intuitions while preserving realism about causation.

Causal relativism has other ramifications that deserve further discussion. For example, the modal profile will play a substantive role in filling in the content of *negations*. For **Alice**, 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

[17] While normally run together, objectivity and realism are substantively different. See (Clarke-Doane 2020, 27) for further elucidation of the distinction.

– 'not-red' refers to all non-red colors. Relativity to modal profile will dictate answers to questions surrounding causation by omission.

§8 References

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

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.

Briggs, Rachael. 2012. "Interventionist Counterfactuals." *Philosophical Studies* 160 (1): 139–66.

Cartwright, Nancy. 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.

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.

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." In *Causality, Probability, and Heuristics: A Tribute to Judea Pearl*, 383–406. London: College Publications.

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

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

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.

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

Hitchcock, Christopher. 2001. "The Intransitivity of Causation Revealed in Equations and Graphs." *The Journal of Philosophy* 98 (6): 273–99.

———. 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.

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

Kim, Jaegwon. 1974. "Noncausal Connections." *Noûs* 8 (1): 41–52.

Lewis, David. 1973. "Causation." *The Journal of Philosophy* 70 (17): 556.

———. 2000. "Causation as Influence." *The Journal of Philosophy* 97 (4): 182.

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.

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

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, UK ; NY: Cambridge University Press.

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

Schroeter, Laura. 2019. “Two-Dimensional Semantics.” Edited by Edward Zalta. *Stanford Encyclopedia of Philosophy*.

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*. NY: Oxford University Press.

———. 2016. “The Problem of Variable Choice.” *Synthese* 193 (4): 1047–72.