Synthese (2023) 201:197
https://doi.org/10.1007/s11229-023-04101-w

ORIGINAL RESEARCH

Conjoined cases

Tomasz Wysocki[1]

Received: 12 April 2022 / Accepted: 9 February 2023 / Published online: 31 May 2023

Abstract
Incorporating normality ascriptions into counterfactual theories of causation was supposed to handle isomorphs. It doesn't—conjoining isomorphs can produce cases that such ascriptions cannot resolve.

Keywords Causation · Isomorphs · Normality · Defaults · Structural equations

1 Introduction

The spectre of isomorphs haunts counterfactual theories of causation, and the standard attempt to dispel it involves employing normality evaluations. This attempt founders.

Counterfactual theories of deterministic causation (Glymour & Wimberly, 2007; Halpern & Pearl, 2005; Hitchcock, 2001; Weslake, 2015; Woodward, 2003) analyze causal claims in terms of counterfactuals (Sect. 2). These theories face the problem of isomorphs (Sect. 3): two cases, called *isomorphs*, share the counterfactual structure, although only one case exemplifies causation (Hall, 2007; Hiddleston, 2005); therefore, counterfactual theories are bound to misjudge either isomorph. To break the isomorphism, some have proposed (call them) *normality theories*, which take into account also normality evaluations of the events involved (Gallow, 2021; Hall, 2007; Halpern, 2016; Halpern & Hitchcock, 2015; Hitchcock, 2007). the theories can distinguish between isomorphs by ascribing different normality evaluations to otherwise analogous events.

This strategy, I argue, fails. The two classic pairs of isomorphs are overdetermination–bogus prevention and early preemption–a short circuit. From either pair, I construct a new *conjoined case* so that however you assign normality to the events, normality theories misjudge one of the causal claims true of the case. Crucially, the judgments lost are the ones that normality evaluations were supposed to save. I first focus on Halpern's latest theory (2016), the currently most popular normality theory

✉ Tomasz Wysocki
tomwysocki@pitt.edu

[1] History and Philosophy of Science, University of Pittsburgh, 1101 Cathedral of Learning, 4200 Fifth Avenue, Pittsburgh, PA 15260, USA

(Sect. 4), but I then adapt my counterexamples to other normality theories too (Sect. 5). I also respond to possible objections (Sect. 6): that I ought not to conjoin cases in the first place, that my models of the cases are inapt, that my criticism is old news, and that I unjustifiably assume that the correct judgments in the conjoined cases must match the judgments in the original cases.

2 Causal models and counterfactual theories

Theories of causation are typically evaluated against how well they predict causal judgments that they cannot convincingly explain away (Paul & Hall, 2013). That's why much energy in this literature has been spent on formulating cases that support one theory over its competitors. With respect to this measure, counterfactual theories (Glymour & Wimberly, 2007; Halpern & Pearl, 2005; Hitchcock, 2001; Weslake, 2015; Woodward, 2003) and normality theories (Gallow, 2021; Hall, 2007; Halpern, 2016; Halpern & Hitchcock, 2015; Hitchcock, 2007) have been most successful.

Counterfactual theories are formulated within the framework of causal models (Pearl, 2000; Spirtes et al., 2000). A *causal model* of a case consists of variables, their ranges, and structural equations; I'll use $\mathfrak{M}$ to denote the model of the case under consideration. Expressions of the form $X{=}x$, where X is a variable and x is a value from X's range, denote *atomic events*. Any non-atomic event is denoted by a boolean combination of expressions denoting atomic events. I'll often use *conjunctive events*, which are conjunctions of atomic events variables; I'll sometimes denote such events with expressions of the form $\vec{X}{=}\vec{x}$, where $\vec{X}$ is a list of distinct variables, and $\vec{x}$ is a list of their corresponding values. The value of every variable is determined by the variable's equation and the values of the variable's *parents*—variables that figure in the equation as arguments. An exogenous equation has no arguments and thus ascribes to the variable its actual value. An endogenous equation has a non-zero arity and determines what happens from what has already happened. An *ancestor* of a variable is a parent of that variable or an ancestor of some parent of that variable; no other variables count as ancestors. An *assignment* is a function that assigns to every variable a value from the variable's range. A *solution* to a model is an assignment that satisfies the model's structural equations. I'll visualize models using diagrams: nodes correspond to variables and hence I'll use these terms interchangeably, values in nodes denote the variables' actual values, and edges pointed at a node come from its parents. I'll deal with acyclic models only, in which no node is its own ancestor. Acyclic models always have a single solution.

For an illustration, take *overdetermination*:

[1] The revolutionary and the spy don't know about each other. It's high time, both think. They poison the tsar's tea. He dies.

a. The revolutionary poisoning the tea rather than retreating caused the tsar to die rather than survive.

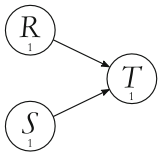

Fig. 1 Overdetermination

The standard representation of overdetermination is this. As atomic events take: $R=1$ if the revolutionary poisons the tea, $R=0$ if he retreats; $S=1$ if the spy poisons the tea, $S=0$ if she retreats; $T=1$ if the tsar dies, $T=0$ if he survives (Fig. 1). Three equations determine what happens:

$$R \leftarrow 1, \quad S \leftarrow 1, \quad T \leftarrow R \vee S \tag{1}$$

The first equation states that the revolutionary poisons the tea. The second, that the spy poisons the tea. The third, that the tsar dies if the tea has been poisoned at least once. The equations yield a solution: $R=S=T=1$.

The framework affords an interventionist semantics of counterfactuals. Where $\mathfrak{M}$ is a model, $\mathfrak{M}_{\vec{X}=\vec{x}}$ is the post-intervention model in which conjunctive event $\vec{X}=\vec{x}$ has been brought about with an intervention. To produce $\mathfrak{M}_{\vec{X}=\vec{x}}$ from $\mathfrak{M}$, for every variable X from $\vec{X}$, replace its equation with $X \leftarrow x$, where x is X's value from $\vec{x}$. The (would-) *counterfactual* "had $\vec{X}=\vec{x}$ happened, φ would have happened," where φ is a (possibly non-atomic) event, holds in $\mathfrak{M}$ *iff* φ holds in (i.e., is satisfied by the solution to) $\mathfrak{M}_{\vec{X}=\vec{x}}$. So, a counterfactual is true *iff* bringing about the antecedent makes the consequent true. For instance, "had the spy failed to poison the tea, the tsar still would have died" holds because $T=1$ on the solution to $\mathfrak{M}_{S=0}$, the model where S's equation has been replaced with $S \leftarrow 0$: the revolutionary's revolutionary act suffices to kill the tsar.

Counterfactuals underlie *counterfactual theories* of causation (Halpern, 2016, Halpern & Pearl, 2005; Hitchcock, 2001; Weslake, 2015; Woodward, 2003). What's crucial about these theories is that, regardless how complicated they make causal judgments based solely on the causal model of the target situation. I'll focus on a particular theory, the contrastive version of Halpern's (2016) *modified theory*, as it underlies Halpern's most recent normality theory, the main target of my current argument. I'll work with the theory's contrastive version, on which the full logical form of a causal claim is "a cause rather than its contrast caused the effect rather than its contrast," as in [1a]. The reasons for this choice are that contrastive claims are more informative and evoke less ambiguous intuitions than non-contrastive claims, and that every non-contrastive interventionist causal theory parasitizes on a contrastive theory. (By the latter I mean that any non-contrastive claim "$\vec{C}=\vec{c}$ causes $E=e$" is translated as a contrastive claim "$\vec{C}=\vec{c}$ rather than $\vec{C}=\vec{c}$ causes $E=e$ rather than $E=\underline{e}$" for some contrast cause $\vec{C}=\vec{c}$ and contrast effect $E=\underline{e}$.) For these reasons, contrastivism has been increasingly embraced in the literature (Gallow, 2021; Maslen, 2004; Northcott, 2008; Schaffer, 2005); even those who seem concerned with a non-contrastive theory do find it necessary to provide a contrastive version (Halpern & Pearl, 2005; Fenton-

Glynn 2017). However, I'll also show (Sect. 5) how to adapt my counterexamples to non-contrastive theories.

On Halpern's contrastive counterfactual theory, a cause rather than its contrast causes an effect rather than its contrast *iff* both the cause and the effect occur, no part of the cause is redundant, and were the contrast cause to occur (while, potentially, some other variables were kept at their actual values), the contrast effect would occur. It's a mouthful; I'll express it formally with three conditions. $\vec{C}{=}\vec{c}$ rather than $\vec{C}{=}\underline{\vec{c}}$ *collectively causes* $E{=}e$ rather than $E{=}\underline{e}$ in a model *iff* there's a (possibly empty) list of *contingency variables* $\vec{T}$ distinct from $\vec{C}$ and E such that

act	$\vec{C}{=}\vec{c} \wedge E{=}e$ happened, and $\vec{C}{=}\underline{\vec{c}} \wedge E{=}\underline{e}$ didn't,
wit	had $\vec{C}{=}\underline{\vec{c}} \wedge \vec{T}{=}\vec{@}$ happened, $E{=}\underline{e}$ would have happened instead of $E{=}e$, i.e., $E{=}\underline{e}$ happens in the witness model $\mathfrak{M}_{\vec{C}=\underline{\vec{c}}\wedge\vec{T}=\vec{@}}$,
min	the definition isn't satisfied for any proper subset of $\vec{C}$.

where $\vec{@}$ denotes $\vec{T}$'s actual values.[1] I'll call $\mathfrak{M}_{\vec{C}=\underline{\vec{c}}\wedge\vec{T}=\vec{@}}$ the *witness* for the target causal claim, for this is the model that shows, together with *act*, the counterfactual dependence of $E{=}e$ on $\vec{C}{=}\vec{c}$.

The theory essentially requires that bringing about the contrast cause under some actual contingency prevents the effect. Now, in a case like overdetermination (to be analyzed shortly), the definition entails that the atomic events that are standardly classified as causes aren't collective causes, but that their conjunction is. Therefore, I'll say that $C{=}c$ rather than $C{=}\underline{c}$ *causes* $E{=}e$ rather than $E{=}\underline{e}$ *iff* $\vec{C}{=}\vec{c}$ rather than $\vec{C}{=}\underline{\vec{c}}$ collectively causes $E{=}e$ rather than $E{=}\underline{e}$, where C is among $\vec{C}$, c is C's value in $\vec{c}$, and $\underline{c}$ is C's value in $\underline{\vec{c}}$ (Halpern, 2016, p. 29).

The theory correctly evaluates the overdetermination case [1]: the revolutionary poisoning the tea ($R{=}1$) rather than retreating ($R{=}0$) caused the tsar to die ($T{=}1$) rather than survive ($T{=}0$). First, consider whether $R{=}1 \wedge S{=}1$ rather than $R{=}0 \wedge S{=}0$ collectively cause $T{=}1$ rather than $T{=}0$. They do. Take $\mathfrak{M}_{R=0\wedge S=0}$ as the witness. Per *act*, the cause and effect happen, and their contrast don't. Per *wit*, were the contrast cause to happen, the contrast effect would too; the contingency here is empty. Per *min*, neither $R{=}1$ rather than $R{=}0$ alone nor $S{=}1$ rather than $S{=}0$ alone collectively causes $T{=}1$ rather than $T{=}0$. What follows next is that $R{=}1$ rather than $R{=}0$ causes *simpliciter* $T{=}1$ rather than $T{=}0$. The theory delivers the correct result.

3 Isomorphism and normality

Any counterfactual theory faces the problem of isomorphs (Hall, 2007; Hiddleston, 2005): there are cases that share the counterfactual structure—i.e., they can be represented with the same model—but elicit different intuitions. Since a counterfactual

[1] For simplicity, I limit the account to atomic effects, whereas Halpern (2016, pp. 25, 81) allows for non-atomic effects.

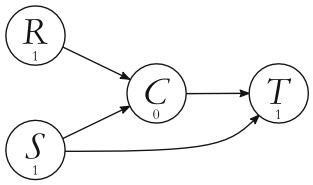

Fig. 2 Bogus prevention with an extra node

theory yields a causal judgment based only on the model of the target situation, no such theory can save both intuitions.[2]

Consider *bogus prevention* (Hiddleston, 2005; Hitchcock, 2007):

[2] The spy has a change of mind: she won't poison the tsarina's coffee. The tsarina has a good guard, very careful. The guard adds an antidote to the tsarina's harmless coffee. The tsarina survives.

a. The guard adding the antidote to the coffee rather than standing back *does not* cause the tsarina to survive rather than die.

For reasons that will become clear momentarily, I'll represent the events rather inconveniently: $R{=}1$ if the guard adds the antidote to the coffee, $R{=}0$ if he stands back; $S{=}1$ if the spy does *not* poison the coffee, $S{=}0$ if she does; $T{=}1$ if the tsarina survives, $T{=}0$ if she dies. Three equations represent the plot:

$$R \leftarrow 1, \quad S \leftarrow 1, \quad T \leftarrow R \vee S. \tag{2}$$

The last equation states that the tsarina will survive if the guard adds the antidote or the spy decides not to poison the coffee.

Never mind what Halpern's, or any counterfactual theory for that matter, says about this case—just compare the model with the one for overdetermination (Fig. 1). Their variables, their ranges, and their equations agree. Yet, the revolutionary lacing the tea ($R{=}1$) rather than retreating ($R{=}0$) *does* cause the tsar to die ($T{=}1$) in [1], while the guard adding the antidote to the coffee (again, $R{=}1$) rather than standing back ($R{=}0$) *does not* cause the tsarina to survive (again, $T{=}1$) rather than die ($T{=}0$) in [2]. You have isomorphic models but opposing judgments. Thus, relying on structural equations alone cannot do justice to both cases—provided they are correctly represented by the models.

To save counterfactual theories, you can deny that equations (2) do justice to the counterfactual structure of the case. In this vein, Blanchard and Schaffer (2017) and McDonald (2023) claim the model misrepresents [2] because a crucial variable is omitted: one for whether the antidote counteracts the poison. Add the variable, and the judgment is accounted for (Fig. 2).

Let $C{=}1$ mean that the guard's antidote counteracts the poison, and $C{=}0$ that it doesn't. Since the antidote neutralizes the poison only if the guard adds it to the coffee

[2] Some authors see the role of intuitions as much less central for the project of defining causation (Beckers & Vennekens, 2018; Clarke, 2023; Glymour et al., 2010; Woodward, 2021); they, of course, may as well welcome a solution that sacrifices one of the intuitions, if this sacrifice is made for good reasons.

while there is poison to neutralize, and the tsarina survives if the spy fails to poison the coffee or the poison is neutralized, the equations read:

$$R \leftarrow 1, \quad S \leftarrow 1, \quad C \leftarrow R \wedge \neg S, \quad T \leftarrow C \vee S. \tag{3}$$

With the updated model, the *act-min* theory solves the case correctly. First, $S{=}1$ rather than $S{=}0$ collectively causes, and hence causes *simpliciter*, $T{=}1$ rather than $T{=}0$. Take $\mathfrak{M}_{S=0 \wedge C=0}$ as the witness. *Act* and *min* are satisfied; *wit* is satisfied because $T{=}0$ on this model. Second, $R{=}1$ rather than $R{=}0$ doesn't collectively cause $T{=}1$ rather than $T{=}0$. Under no freezing C or S at their actual values does $R{=}0$ entail $T{=}0$, which means that there's no witness for the claim. Third, $S{=}1 \wedge R{=}1$ rather than $S{=}0 \wedge R{=}0$ doesn't collectively cause $T{=}1$ rather than $T{=}0$ because the claim violates *min*, as $S{=}1$ rather than $S{=}0$ alone collectively causes $T{=}1$ rather than $T{=}0$. Therefore, *act-min* correctly predicts that the guard adding the antidote doesn't cause the tsarina to survive.

Another solution— the focus of my argument—is to extend counterfactual theories with normality evaluations (Hall, 2007; Halpern, 2016; Halpern & Hitchcock, 2015; Hitchcock, 2007; Menzies, 2017).[3] Implementations differ, but almost all such theories stem from the same principle: the witness must be at least as normal as the actual solution (where the opposite of 'at least as normal' is 'less normal or incomparable').[4]

To incorporate the distinction, Halpern adds a fourth condition:

nrm the witness model from *wit* is at least as normal as the actual situation, $\mathfrak{M}_{\vec{C}=\vec{c}\wedge\vec{T}=\vec{t}@} \succcurlyeq \mathfrak{M}$,

where $\mathfrak{M} \succcurlyeq \mathfrak{m}$ means that $\mathfrak{M}$ is at least as normal as $\mathfrak{m}$; the order on models fully depends on the order of their solutions, i.e., if two models have the same solution, they are equally normal, $\mathfrak{M} \approx \mathfrak{m}$. With the extra condition, the theory states that bringing about the contrast cause under some actual contingency prevents the effect *without making the situation any less normal*. Mnemonically: causation requires a witness no less normal than the actual situation. I'll focus on Halpern's theory, which I'll call *act-nrm*, for it seems to be the most popular normality theory out there. I'll discuss other theories later, however (Sect. 5).

The normality order is partial—reflexive, (weakly) asymmetric, and transitive—but it needn't be total, i.e., not every two models must be comparable.[5] Normality theories typically assume that comparing two models reduces to comparing variable values between the solutions to these models; that means that variables' ranges come ordered. So, for models $\mathfrak{M}$ and $\mathfrak{m}$ that share variables, if every value in $\mathfrak{M}$'s solution is at least as normal as its counterpart in $\mathfrak{m}$'s solution, $\mathfrak{M} \succcurlyeq \mathfrak{m}$; if some of the values

[3] On yet another one (Beckers and Vennekens, 2018), causal models are extended with event timing; causal claims depend on both the casual structure of the case and the timing of the events involved. I'll leave engaging with that proposal for another occasion; I'll just flag that this strategy is similar to the one from normality in that on both, causal claims hold in virtue on some properties of events beyond the counterfactual structure.

[4] In Sect. 5, I present two theories that use the normality of events but not solutions, Harinen's (2017) and Gallow's (2021).

[5] $\approx$ is the symmetric part of $\succcurlyeq$.

in $\mathfrak{M}$'s solution are strictly more normal and some strictly less normal than their counterparts from $\mathfrak{m}$'s solution, the models are incomparable; the models are also incomparable if some value in $\mathfrak{M}$'s solution and its counterpart in $\mathfrak{m}$'s solution are incomparable. However, sometimes this assumption is lifted. Halpern and Hitchcock (2005) allow that if some values in a single solution denote events that rarely occur together, the solution can be deemed less normal than, say, the actual solution, even though pairwise comparing the values between the solutions would deem both models equally normal. They also assume that the outcomes of structural equations, when fed actual values, are normal, though they treat this rule as defeasible. The details depend on the theory. I intend my counterexamples to work even against theories that don't reduce comparing models to comparing values.

The concept of normality, as applied to events, is supposed to be (or at least to explicate) the everyday concept of normality. For McGrath (2005, p. 138; see also Thomson, 2003, p. 100), who uses the concept to investigate causal omissions, a behavior is normal for an object *iff* the object is supposed to display this behavior. What an object is supposed to do, says McGrath, depends on the category of the object: artifacts have intended functions, organs have normal functions, organisms have natural behaviors, moral agents have norms to follow, and so on. Halpern and Hitchcock (2010, p. 402, 2015, pp. 431–432) concur. The normality of an event, they say, depends on the frequency of the type of the event in question, moral and social norms, and norms of proper functioning.[6] Others speak of default (rather than normal) events—the events that would have transpired if nothing had affected the system from without (Hall, 2007; Hitchcock, 2007; Maudlin, 2004; Menzies, 2004; Wolff, 2016).[7]

Normality theories can distinguish between overdetermination and bogus prevention. In [1], assume that poisoning the tea is less normal than refraining from poisoning the tea, and therefore $\mathfrak{M}_{R=0\wedge S=0} \succcurlyeq \mathfrak{M}$. Hence, per *nrm*, $\mathfrak{M}_{R=0\wedge S=0}$ makes for a good witness. Since the remaining conditions still hold, the theory yields that the revolutionary poisoning the tea rather than retreating caused the tsar to die rather than survive. Assume the same for [2]: poisoning the coffee is less normal than standing back. But that means that a model where the spy poisons the coffee is less normal than or incomparable to a model where she doesn't. To derive that $R{=}1$ rather than $R{=}0$ causes $T{=}1$ rather than $T{=}0$, you need $\mathfrak{M}_{R=0\wedge S=0}$ as a witness, but this time $\mathfrak{M}_{R=0\wedge S=0}$ isn't at least as normal as $\mathfrak{M}$. The theory correctly rules that the guard adding the antidote rather than standing back didn't cause the tsarina to survive rather than die.

4 Conjoined cases

Yet, the problem of isomorphs resurfaces as the problem of *conjoined cases*. Say that *act-nrm* solves an instance of the problem of isomorphs by assigning different normality evaluations to the values of some variables. Conjoin the two cases by these variables (and possibly others) in such a way that any witness that allows the theory

[6] This also seems to match how the folk make normality ascriptions (Bear & Knobe, 2017; Hitchcock & Knobe, 2009; Wysocki, 2020).

[7] Halpern and Hitchcock (2015) call witnesses (ab)normal and events (a)typical. I use a single term for both.

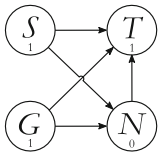

Fig. 3 The Tsarina. (Overdetermination and bogus prevention conjoined)

to vindicate one intuition forces the theory to contradict the other intuition. You've produced a *conjoined case*—a counterexample to the theory. I'll use two classic pairs of isomorphs—overdetermination and bogus prevention (explained above) and early preemption and a short circuit—to produce two conjoined cases; one advantage of this strategy is that I'll reuse models of the conjoined cases that proponents of normality theories espouse themselves. Some remarks: first, while in what follows I conjoin a case and its isomorph, you in fact only need to conjoin cases that require different normality evaluations of the same event; the cases needn't be counterfactually isomorphic. Second, I enrich the counterfactual structure of the cases so that no normality ordering can preserve the judgments.

As the first conjoined case, take *The Tsarina*:

[3] Tea and goose will be served for dinner. The spy's mission is to kill one of the royals; she chooses the tsar and poisons his tea.
The tsarina has a loyal guard and an ulterior agenda. The guard can add ginseng to the goose. The herb works differently in either royal: it protects the tsarina from poison but is lethal to the tsar.
The last detail: the tsar suspects the tsarina is scheming, and he expects the spy to poison her this morning. If the tsarina does not die during dinner, he will realize he has been foiled and is in grave danger, and his heart will give out.
The royals enjoy the goose. The tsar has his tea. The tsarina survives. He dies in terror.

a. The spy poisoning the tsar rather than the tsarina caused the tsar to die rather than survive.
b. The guard adding ginseng to the goose rather than standing back didn't cause the tsarina to survive rather than die.

As atomic events, take: $S{=}1$ if the spy poisons the tsar's cup, $S{=}\text{-}1$ if the tsarina's; $S{=}0$ if none; $G{=}1$ if the guard adds ginseng to the goose, $G{=}0$ if not; $T{=}1$ if the tsar dies, $T{=}0$ if not; $N{=}1$ if the tsarina dies, $N{=}0$ if not. Equations

$$S \leftarrow 1, \quad G \leftarrow 1, \quad N \leftarrow S{=}\text{-}1 \wedge G{=}0, \quad T \leftarrow S{=}1 \vee G{=}1 \vee N{=}0 \quad (4)$$

represent the counterfactual structure of the case and yield actual values $S{=}G{=}T{=}1$ and $N{=}0$ (Fig. 3).

The case conjoins overdetermination (the STG part of the model) with bogus prevention (SNG) and is analyzed on *act-min* as follows. Saving [3a] requires that $S{=}1 \wedge G{=}1$ rather than $S{=}\text{-}1 \wedge G{=}0$ collectively cause $T{=}1$ rather than $T{=}0$, which

entails that S=1 rather than S=-1 causes T=1 rather than T=0. There is only one model that can serve as a witness for the claim, $\mathfrak{M}_{S=\text{-}1\wedge G=0}$, for only on this witness do the contrast cause and effect happen, S=-1 $\wedge$ T=0. Therefore, $\mathfrak{M}_{S=\text{-}1\wedge G=0}$ needs to be at least as normal as $\mathfrak{M}$, the model of the actual situation. But if so, then the theory fails to predict [3b]. On $\mathfrak{M}_{S=\text{-}1\wedge G=0}$, N=1 happens, which means the model is also a witness for S=1 $\wedge$ G=1 rather than S=-1 $\wedge$ G=0 collectively causing N=0 rather than N=1, and hence for G=1 rather than G=0 causing N=0 rather than N=1. Handling [3a] entails mishandling [3b].[8]

I can now explain two details of the case. First, T needs to depend on N counterfactually—if it didn't, $\mathfrak{M}_{S=\text{-}1\wedge G=0\wedge N=0}$ could also serve as a witness for [3a], but this one doesn't contradict [3b]. With the extra arrow from N to T, you cannot freeze N at its actual value, for T won't budge. Whence I had the tsar's death counterfactually depend also on the tsarina surviving. This is a move I'll use in the other case too.[9] Second, I stipulate that the poison in the cup, ginseng in the goose, and the terror of seeing the tsarina alive are on par, i.e., none of them preempts the other ones.

The other counterexample conjoins early preemption (Hitchcock, 2001) with a short circuit (Hall, 2007). First consider the original cases. *Early preemption*: the quisling poisons the tsar's soup, but if he hadn't, the spy would have. The quisling but not the spy caused the tsar's death. *Short circuit*: the spy laces the patriarch's soup with poison but only if the quisling adds an antidote first. The patriarch survives, but the quisling didn't cause it. The cases share the counterfactual structure: the spy's action depends on the quisling's, and the actions together determine what happens to the third character in the case.

On the standard solution from normality, in early preemption, treat the model where neither the quisling nor the spy poison the soup as more normal than the actual situation, for murder attempts violate norms of morality and good taste. Use this model as a witness and you'll get the correct causal claim. In the short circuit, where the spy's actions depend on the quisling's, not poisoning the soup is more normal than poisoning the soup if there's already an antidote in it, which is still more normal than poisoning the soup even if there's no antidote (Halpern & Hitchcock 2015:451). Halpern and Hitchcock justify this ordering: not poisoning is the most normal because it violates no norms; ineffective poisoning follows because it's deceitful but not murderous; effective poisoning is murderous and therefore the most abnormal. Therefore, a model in which the quisling doesn't add the antidote but the spy does add the poison is less normal than the actual model and therefore cannot serve as a witness for the quisling causing the tsar to survive.

With this in mind, consider *The Patriarch*:

[4] The tsar has ruled together with the patriarch. But not for long. The spy and the quisling are set to kill him. The tsar is lethally allergic to tarragon and the

[8] *Act-nrm* also mistakenly denies that N=0 rather than N=1 causes T=1 rather than T=0; this problem has been already pointed out by Rosenberg and Glymour (2018). However, the current counterexample doesn't piggyback on that previous problem. Say, you could use $\mathfrak{M}_{S=\text{-}1\wedge G=0\wedge N=0}$ as a witness to vindicate that N=0 rather than N=1 causes T=1 rather than T=0. Because $\mathfrak{M}_{S=\text{-}1\wedge G=0\wedge N=0}$ shares the solution with $\mathfrak{M}_{S=\text{-}1\wedge G=0}$, they are equally normal, and the latter could be used as a witness to contradict [3b].

[9] I am indebted to Sander Beckers for suggesting this move.

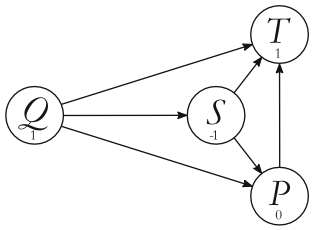

Fig. 4 The Patriarch. (Early prevention and a short circuit conjoined)

patriarch to paprika. But if the patriarch eats paprika with tarragon, he will recover—tarragon will cure him.
The tsar and the patriarch feast together. Soup will be served. The quisling can spice it with tarragon or retreat. If the quisling goes through, the spy will add paprika. Otherwise, she will spice the soup with tarragon. Maybe to remove suspicion from the patriarch, who planned the murder? No one can tell.
The tsar has suspected the patriarch and expects the spy to poison him. If the patriarch does not die during the feast, the tsar will realize he has been outsmarted, and his heart will give out. The quisling adds tarragon to the soup. The spy adds paprika. This is the last feast for the tsar. But not for the patriarch—he lives on.

a. The quisling lacing the soup rather than retreating caused the tsar to die rather than survive,
b. but it didn't cause the patriarch to survive rather than die.

Atomic events: Q=1 if the quisling poisons the soup with tarragon, Q=-1 if not; S=1 if the spy adds tarragon, S=-1 if paprika, S=0 if nothing; T=1 if the tsar dies, T=0 if he survives; P=1 if the patriarch dies, P=0 if he survives (Fig. 4). Equations

$$Q \leftarrow 1, \quad S \leftarrow \text{-}Q, \quad P \leftarrow Q{=}\text{-}1 \wedge S{=}\text{-}1, \quad T \leftarrow Q{=}1 \vee S{=}1 \vee P{=}0 \qquad (5)$$

yield that Q=T=1, S=-1, and P=0. QST is the early-preemption part of the model; QSP is the short-circuit part.

For Q=1 rather than Q=-1 to cause T=1 rather than T=0, after bringing about the contrast cause Q=-1, you need to to leave P unfrozen, so that it can change its value to P=1, for otherwise the tsar still would die from realizing that he has been foiled. You also need to freeze S at -1, for otherwise the patriarch wouldn't die, and hence the tsar still would. Indeed, on $\mathfrak{M}_{Q=\text{-}1 \wedge S=\text{-}1}$, the tsar survives, T=0, as required by *wit*. Since no other model can serve as a witness, it must be that $\mathfrak{M}_{Q=\text{-}1 \wedge S=\text{-}1} \succcurlyeq \mathfrak{M}$ to satisfy *nrm* and vindicate [4a]. However, the same model can now serve as a witness for the quisling spicing the soup with tarragon (rather than not) causing the patriarch to survive (rather than die), as P=0 on this model. Halpern's theory cannot save [4a] and [4b] at once.

The strategy from conjoined cases in a single albeit lengthy sentence: a conjoined case works because the witness for predicting that the causal relation holds where it should also serves as a witness for predicting that the causal relation holds where it shouldn't.

5 Other normality theories

The counterexamples work against Halpern's newest theory; however, there are other interventionist theories out there, both contrastive and non-contrastive. Conjoining would be more effective a strategy if it worked against these theories too. And it often does.

First, consider the contrastive version of the predecessor of *adm-nrm* (Halpern & Hitchcock, 2015), which adds a normality condition to what Halpern and Pearl (2005) called their *updated theory*. For simplicity, I'll present the theory for atomic causes and effects. $C{=}c$ rather than $C{=}\underline{c}$ causes $E{=}e$ rather than $E{=}\underline{e}$ *iff* all variables can be partitioned into $\{C, E\}$, mediating variables $\vec{M}$, and contingency variables $\vec{T}$, such that for some $\vec{t}$,

U1 $C{=}c \wedge E{=}e$ happened, and $C{=}\underline{c} \wedge E{=}\underline{e}$ didn't,
U2 had $C{=}\underline{c} \wedge \vec{T}{=}\vec{t}$ happened, $E{=}\underline{e}$ would have happened,
U3 for every subset $\vec{M}'$ of mediating variables $\vec{M}$, and every subset $\vec{T}'$ of contingency variables $\vec{T}$, had $C{=}c \wedge \vec{M}'{=}\vec{@} \wedge \vec{T}'{=}\vec{t}$ happened, $E{=}e$ would have happened,
U4 the witness model is at least as normal as the actual model, $\mathfrak{M}_{C=\underline{c}\wedge\vec{T}=\vec{t}} \succcurlyeq \mathfrak{M}$.

Unlike in *adm-nrm*, causation *simpliciter* is now defined directly rather than through collective causation. The role of mediating variables $\vec{M}$ is to mediate the counterfactual dependence of $E{=}e$ on $C{=}c$; the role of contingency variables $\vec{T}$ is to expose this counterfactual dependence, which may be masked by other causes of $E{=}e$. *U2* states that under the contingency $\vec{T}{=}\vec{t}$, the contrast cause prevents the actual effect. *U3* states that if you freeze any number of the mediating variables at their actual values and any number of the contingency variables at their contingency values, the actual cause will still entail the actual effect.[10] The role of this constraint is to limit admissible contingencies to the ones that leave the causal process from C to E intact; if a contingency modified the process, it could introduce what would look like a genuine counterfactual dependence of $E{=}e$ on $C{=}c$. As for *U4*, as before, two models are equally normal if they share the solution.

As Halpern and Hitchcock's (2015) illustrate this theory sufficiently, I'll go straight for the counterexample, The Tsarina (Fig. 3). To get the claim that $S{=}1$ rather than $S{=}\text{-}1$ causes $T{=}1$ rather than $T{=}0$, you must use $\mathfrak{M}_{S=\text{-}1\wedge G=0}$ or $\mathfrak{M}_{S=\text{-}1\wedge G=0\wedge N=1}$ as a witness, for only on these two does $T{=}0$. Say, you use the first witness. *U4* entails that $\mathfrak{M}_{S=\text{-}1\wedge G=0} \succcurlyeq \mathfrak{M}$. But since $N{=}1$ on $\mathfrak{M}_{S=\text{-}1\wedge G=0}$, you can use this witness to arrive at $G{=}1$ rather than $G{=}0$ causing $N{=}0$ rather than $N{=}1$. Say, in turn, you use $\mathfrak{M}_{S=\text{-}1\wedge G=0\wedge N=1}$ as a witness. Per *U4*, $\mathfrak{M}_{S=\text{-}1\wedge G=0\wedge N=1} \succcurlyeq \mathfrak{M}$. But since $\mathfrak{M}_{S=\text{-}1\wedge G=0\wedge N=1}$ and $\mathfrak{M}_{S=\text{-}1\wedge G=0}$ share the solution, $\mathfrak{M}_{S=\text{-}1\wedge G=0\wedge N=1} \approx \mathfrak{M}_{S=\text{-}1\wedge G=0}$, and by transitivity, $\mathfrak{M}_{S=\text{-}1\wedge G=0} \succcurlyeq \mathfrak{M}$. Again, you can use $\mathfrak{M}_{S=\text{-}1\wedge G=0}$ as a witness for the claim that $G{=}1$ rather than $G{=}0$ causes $N{=}0$ rather than $N{=}1$. Saving [3a] means losing [3b].[11]

[10] In $\vec{M}'{=}\vec{@}$ and $\vec{T}'{=}\vec{t}$, I abused the notation; read the expressions as if the assignment on the right side were trimmed to the values of the variables on the left side.

[11] The theory still can handle The Patriarch in their current formulation, but not [6], a non-contrastive version of The Patriarch.

Another recent normality theory is by Gallow (2021). I won't present it in its entirety, for the theory is rather complex. What's relevant here, though, is that if the cause variable parents the effect variable, the theory requires that the effect counterfactually depends on the cause while the contrast cause is normal and the actual cause deviant; *or* the contrast effect is normal and the actual effect deviant; *or* both the cause and the effect are normal and their contrasts are deviant. Which of the three conditions must be satisfied is a free parameter of the theory, i.e., Gallow in fact proposes three separate theories. His theory differs from standard normality theories in two respects: it employs a binary distinction between normal and deviant events (rather than the usual more-normal-than relation), and the distinction applies only to variable values (rather than entire models).

Take first the incarnation of Gallow's theory on which causes must be more normal than their contrasts. If in The Tsarina you deem $G=1$ as deviant and $G=0$ as normal, the theory predicts correctly that the guard caused the tsar to die, [3a], but incorrectly that the guard caused the tsarina to survive, [3b]. Reverse the classification or give both events the same status, and you'll save [3b] but sacrifice [3a].

If effects must be more normal than their contrasts, you'll lose the judgments in bogus prevention. If $N=0$ is deviant and $N=1$ is default, the theory incorrectly judges that the guard caused the tsarina to survive. If $N=0$ is default and $N=1$ is deviant (or both events have the same status), the theory incorrectly judges that the spy replacing the poisonous cup didn't cause the tsarina to survive.

If both causes and effects must be more normal than their contrasts, to save the judgment that $G=1$ rather than $G=0$ and $N=0$ rather than $N=1$ cause $T=1$ rather than $T=0$, you must deem these events deviant and $G=0$, $N=1$, and $T=1$ normal. But if so, you'll incorrectly classify $G=1$ rather than $G=0$ as a cause of $N=0$ rather than $N=1$. This last analysis also works against Harinen's (2017) theory. The theory, inspired by research in the cognitive science of causation, requires that normal effects have normal causes, whereas deviant effects have deviant causes. On this theory, to save the judgment that $G=1$ rather than $G=0$ and $N=0$ rather than $N=1$ cause $T=1$ rather than $T=0$, you must give them the same status (deviant or normal) and give $G=0$, $N=1$, and $T=1$ the opposite status (normal or deviant). But on this classification, $G=1$ rather than $G=0$ incorrectly counts as causing $N=0$ rather than $N=1$.

How about non-contrastive theories? A non-contrastive version of *act-nrm* handles The Tsarina. The judgments to be explained now are that the spy poisoning the tsar caused his death, but that the guard adding the antidote to the tea didn't cause the tsarina to survive. Remember that for a non-contrastive claim to hold, a contrastive claim needs to hold for some contrast cause and effect. So, let $\mathfrak{M}_{S=0\wedge G=0} \succcurlyeq \mathfrak{M}$. This will give you that the spy poisoning the tsar rather than doing nothing caused the tsarina to die rather than survive, which in turn entails the first non-contrastive judgment. Also, let $\mathfrak{M}_{S=\text{-}1\wedge G=0}$ be incomparable with $\mathfrak{M}$. This will block the claim that the guard adding the antidote rather than not caused the tsarina to survive rather than not, which in turn saves the second non-contrastive judgment.

Yet, first, Halpern's (2016) non-contrastive theory still fails for The Patriarch. Second, I would be suspicious of a theory that succeeds for non-contrastive judgments but fails for contrastive ones; the latter are no worse—if not better—data points than

the former. Third, there are versions of the cases that work against non-contrastive theories too.

[5] Su and Go are preparing an experiment with two species of bacteria, *Turicella* and *Nocardia*, mixed in the same petri dish. Su is responsible for setting the temperature of the petri dish. If she warms it up, *Turicella* will die; if she doesn't, *Nocardia* will die. Go decides whether to feed the bacteria ginseng; if she does, *Turicella* will die, and *Nocardia* will withstand the cold. Lastly, *Nocardia*, if not killed, will crowd out *Turicella*. Su warms the petri dish up, and Go adds ginseng. *Turicella* slowly disappears. *Nocardia* flourishes.

a. Warming up the petri dish caused *Turicella* to die.
b. Adding ginseng didn't cause *Nocardia* to survive.

All variables are binary, with 1 denoting 'yes' and 0 'no'. Where S stands for whether Su heats up the petri dish, G for whether Go adds ginseng, T for whether *Turicella* dies, and N for whether *Nocardia* dies,

$$S \leftarrow 1, \quad G \leftarrow 1, \quad N \leftarrow \neg S \wedge \neg G, \quad T \leftarrow S \vee G \vee \neg N \tag{6}$$

represent the counterfactual structure of the case; the model is identical to (4) except for S, which is now binary.

Halpern's (2016) theory, i.e., the non-contrastive counterpart of *act-nrm*, cannot account for the case. Analogously to the original analysis of The Tsarina, the only witness that may save [5a] is $\mathfrak{M}_{S=0\wedge G=0}$, but this witness vindicates $G{=}1$ causing $N{=}0$, contradicting [5b]. Halpern and Hitchcock's (2015) theory fails too. You may save [5a] with two witnesses only, $\mathfrak{M}_{S=0\wedge G=0}$ and $\mathfrak{M}_{S=0\wedge G=0\wedge N=1}$, but since they share the solution, $\mathfrak{M}_{S=0\wedge G=0} \approx \mathfrak{M}_{S=0\wedge G=0\wedge N=1}$. Therefore, whichever witness you choose, it follows that $\mathfrak{M}_{S=0\wedge G=0} \succcurlyeq \mathfrak{M}$, implying that $G{=}1$ causes $N{=}0$ on this theory. [5b] is lost.

While Halpern's (2016) theory, in both its contrastive and non-contrastive versions, fails for The Patriarch, Halpern and Hitchcock's (2015) doesn't, for $\mathfrak{M}_{Q=0\wedge S=0\wedge P=1}$ as a witness vindicates [4a] but not [4b]. Here's an analog of the case, which the theory, contrastive or not, cannot handle.

[6] Qi and Su are preparing an experiment with *Turicella* and *Pantoea*. Qi is deciding whether to add tarragon to the petri dish; if she does, *Turicella* will die, and *Pantoea* will be shielded against the cold. Su is responsible for setting the temperature of the petri dish. If she warms it up, *Turicella* will die; if she doesn't, *Pantoea* will die. Su will warm up the petri dish if Qi doesn't add tarragon to it. Lastly, *Pantoea*, if not killed, will crowd out *Turicella*. Qi feeds tarragon to the bacteria; Su doesn't heat the petri dish. *Turicella* gradually disappears. *Pantoea* flourishes.

a. Warming up the petri dish caused *Turicella* to die,
b. but it didn't cause *Pantoea* to survive.

All variables are again binary. Where Q stands for whether Qi adds tarragon to the dish, S for whether Su warms up the petri dish, T for whether *Turicella* dies, and P for whether *Pantoea* dies, the model is:

$$Q \leftarrow 1, \quad S \leftarrow \neg Q, \quad P \leftarrow \neg Q \wedge \neg S, \quad T \leftarrow Q \vee S \vee \neg P. \tag{7}$$

The model is identical to (5) except for binary S.

To save [6a], Halpern and Hitchcock (2015) can choose from two witnesses, $\mathfrak{M}_{Q=0 \wedge S=0}$ or $\mathfrak{M}_{Q=0 \wedge S=0 \wedge P=1}$. As they share the solution, $\mathfrak{M}_{Q=0 \wedge S=0} \approx \mathfrak{M}_{Q=0 \wedge S=0 \wedge P=1}$. Therefore, whichever model serves as a witness for [6a], $\mathfrak{M}_{Q=0 \wedge S=0} \succcurlyeq \mathfrak{M}$. Since $\mathfrak{M}_{Q=0 \wedge S=0}$ can serve as a witness for Q=1 causing P=0, [6b] is lost. In fact, cases [5] and [6] also contradict *act-nrm*, which makes them the most effective thought experiments supporting my argument.

6 The complaints book

No objections are without objections. One: *I shouldn't conjoin the models in the first place*, as they are used for evaluating distinct causal judgments. The objection won't do. I don't conjoin models, I conjoin situations and then model them with a single model. Nothing's wrong with that. On the standard procedure, you represent a situation with a model and then use it to derive any causal judgment that can be formulated with the model's variables. Indeed, a single model can and has been used to yield multiple causal judgments. For instance, in the standard treatment of the early preemption case (the QST part of the model from Fig. 4), the same model is used to arrive at two judgments: that the quisling, but not the spy, caused the tsar to die.

The second objection: although it's fine to model each conjoined case with one model, *the models I use are inapt*. I could account for the intuitions if I used more (or different) variables. When conjoining overdetermination and bogus prevention (Fig. 3), for instance, I could use the extended model for bogus prevention (Fig. 2) instead of the three-node original. This objection won't do. Notice that I'm conjoining models standardly used in the literature to represent these cases. If my model is inapt, then by analogy, so is the original model for bogus prevention (Fig. 1). That is, this objection capitulates to Blanchard and Schaffer's (2017) argument that invoking normality is unnecessary because once you account for all relevant events, isomorphs turn out not to be isomorphic after all.

Now—you may press—even though I build my cases out of other cases, this doesn't automatically mean that the best model of the conjoined cases conjoins the original models. Conjoining cases, that is, might not be so innocent, and might warrant adding new variables. This objection, I think, unduly shifts the burden. If there are better models of these situations that allow normality theories to yield the correct judgment, then it's on the proponent of the objection to produce such models *and* to explain why I can't use the ones I use. The latter postulate is crucial. It's not enough to claim that there are other apt models; one needs to show that I'm mistaken in using mine.

Still, I'll say more. Although there are no uncontroversial sufficient aptness conditions for models, there are uncontroversial necessary conditions (Blanchard & Schaffer, 2017; Halpern & Hitchcock, 2010; Halpern & Pearl, 2005; Hitchcock, 2001; Pearl, 2000; Spirtes et al., 2000).

I'll just list them. The model entails only true counterfactuals. Atomic events denoted by values of different variables aren't conceptually related (e.g., one vari-

able can't represent whether the tsar is alive and another whether he's dead because a situation where he's alive while being dead is incoherent). Different values of the same variable denote inconsistent events (the quisling can't poison and not poison the soup at the same time). Variable values shouldn't represent events that we consider too remote and thus aren't willing to take seriously (e.g., that the spy bilocates and replaces both poisoned cups at the same time). No combination of values of different variables may represent an incoherent state of affairs (Ross & Woodward, 2021).[12] The models that represent the conjoined cases clearly satisfy these criteria.

The third objection: *old news, old news*; all I did was produce more examples, but no one claims that normality solves every problem (Halpern, 2016, p. 90). Just add my cases to the pile to be solved by another improvement to the counterfactual theory. I disagree. The main problem is that my conjoined cases, as I've stressed already, reuse cases that normality was supposed to handle. Imagine that a future iteration of the counterfactual theory accounts for the conjoined cases. When applied to the conjoined cases, this future theory won't appeal to normality (at least in the same way that current normality theories do). But if so, then you should be able to apply the theory to original isomorphs without appealing to normality either, for the models of isomorphs are just sub-models of conjoined cases.

The argument is new in another respect. While I agree with Blanchard and Schaffer (as well as with McDonald, 2023) that incorporating normality into counterfactual theories doesn't work, I'm not so convinced by their positive proposals. In their treatment of bogus prevention (Fig. 2), they add a variable that represents "whether or not any neutralization occurs" (Blanchard & Schaffer, 2017, p. 201). Adding the variable seems to violate one of the aptness criteria, for the event of 'neutralization occurring' seems to require, as a matter of conceptual truth, the neutraliz*er* and the neutraliz*ed*. That is, an assignment where there's no poison ($S=1$) but neutralization occurs ($C=1$) seems incoherent. If the tea isn't poisoned, there's no meaningful way in which the antidote can counteract the nonexistent poison.[13] Therefore, unlike Blanchard and Schaffer, I don't maintain that the standard models of isomorphs are inapt.

The fourth objection: *the correct judgments in the conjoined cases needn't match the judgments in the original cases*. It might be, that is, that conjoining cases warrants a change in causal judgments.

I do have considerable evidence that people—my peers, to be exact—do share the intuitions in the conjoined cases. However, you may also understand the current argument as opening up space for a future experimental philosophy study that tests my cases against laypersons' judgments. If they agree with my evaluations, it would support my case against normality theories. (I must stress, though, that such a study should be run cautiously. Conjoining produces complicated cases, and keeping all relevant details in mind may be hard for a layperson. Therefore, the study would have to ensure that answers are due to the participants' competence rather than a performance error.)

[12] The most convincing reason I see for this requirement is that an intervention cannot bring about a state of affairs that cannot exist, and interventions can bring about any combination of values of different variables.

[13] Gallow (2021), Ross and Woodward (2021), and Hitchcock (2007, p. 527) probably would concur.

There is, however, a complementary interpretation of my argument, on which the intuitive responses to the conjoined cases don't matter much: there's no principled reason to change the judgment between the original and the conjoined cases. Consider, for instance, the overdetermination relation between the tsar's death and the actions of the spy and the guard from The Tsarina. If you're initially told that the tsar will die regardless of whether the guard adds ginseng to the goose, you'll think that the guard didn't cause the tsar's death. However, once you learn that, were the tea not poisoned, the tsar would die if and only if ginseng were in the goose—i.e., that the case is one of overdetermination—you'll revise your judgment. The revision is principled: you simply realize that the counterfactual dependence of the effect on the cause was masked by another cause. In contrast, there's no principled reason to change the evaluation of what causes $T=1$ once you extend the STG model with counterfactual relationships between S, G, and N, and then between N and T. We simply don't have independent instances—other, simpler cases, for example—where this kind of change in the case leads to a change of judgment.

I therefore propose the following. If the conjoined cases elicit unambiguous intuitions that oppose the judgments I assumed, *and* that intuition is explained by a normality theory, then such intuitions would support the supposition that conjoining standalone cases may alter causal relations, and my cases, contrary to my intentions, would bolster the normality approach. If the conjoined cases elicit unambiguous intuitions that match my judgments, my objection to normality theories succeeds. However, if the conjoined cases elicit weak or ambiguous intuitions, my argument still stands *unless* there are principled reasons for changing the judgments between the original and conjoined cases. Given that conjoining cases doesn't introduce elements already known to change causal judgments, I take it there are no such principled reasons.

I'm not alone in proposing this strategy, as a somewhat similar one has been recently used by Clarke (2023). He argues that intuitions should never be used as decisive evidence for a theory of causation because no theory can simultaneously accommodate intuitions in preemption and short-circuit cases. He supports the latter claim by showing that you can extend a preemption case to a more complex case, you can extend a short-circuit to the same complex case, and that neither extension introduces elements that warrant changing the causal judgment. If you treat the final case as a richer version of preemption, a certain event should count as a cause of a certain other event, but if you treat the final case as a richer version of a short-circuit, the first event should *not* count as a cause of the second event. Analogously to my response to the current objection, Clarke's argument doesn't depend on the intuitions that the final case elicits, but on the assumption that extending the cases doesn't introduce elements that alter causal relations.[14]

[14] The difference between my strategy and his is that, unlike Clarke, I am ready to admit that if the final case elicits unambiguous intuitions inconsistent with the original case, I may have a reason to treat the changes as altering causal relations.

7 Conclusion

I have not offered just any counterexamples against current normality theories—the theories were designed, I daresay, to handle these sorts of cases. That's not all. I have also offered a recipe for brewing more problems. Say that some theory handles a particular pair of cases by assigning different normality evaluations to the values of some variables.[15] Conjoin the two cases by (at least) these variables in a way that any witness that allows the theory to vindicate one intuition forces the theory to contradict the other intuition. You've produced a counterexample. I don't claim this will always work, but so far the recipe seems successful.

Now, I'm not saying that normality doesn't play a role in lay causal judgments. The empirical evidence for that is too strong (see, e.g., Alicke et al., 2015; Clarke et al., 2015; Gerstenberg & Icard, 2020; Henne et al., 2017; Henne et al., 2021; Hitchcock & Knobe, 2009; Knobe & Fraser, 2008; Kominsky & Phillips, 2019)[16]. Nor am I saying *here* that normality shouldn't play a role in causal judgments. I am saying, though, that the way current normality theories incorporate normality doesn't work. Assigning normality to variable values overlooks the fact that the same event can be causally efficacious toward one effect and inefficacious toward another (non)effect. Branding events as normal/abnormal or default/deviant simply doesn't leave room for this obvious option. The spectre of isomorphs has yet to meet its exorcist.

Acknowledgements This paper benefited greatly from the feedback of friends and foes alike: Colin Allen (and the feedback group he organized), Dmitri Gallow, JP Gamboa, Eric Hiddleston, Caitlin Mace, Katie Morrow, Jim Woodward, and the audiences at the 2021 PSA and the 2022 Central APA. Above all, I am grateful to the reviewers for their thorough suggestions. One of the reviewers, who later turned out to be Sander Beckers, proposed how to amend my cases so they work against any normality ordering of solutions, not only standard ones. Amending the cases made the argument much stronger, as it now targets any present or future theory that relies on ordering solutions according to their normality. I wish every author a reviewer like Sander. This work was supported by the Chateaubriand Fellowship of the Office for Science & Technology of the Embassy of France in the United States and a Reinhart Koselleck project (WA 621/25-1), funded by the Deutsche Forschungsgemeinschaft (DFG).

Declarations

Conflicts of Interest None declared.

References

Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., & Lagnado, D. A. (2015). Causal conceptions in social explanation and moral evaluation: A historical tour. *Perspectives on Psychological Science, 10*(6), 790–812.

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition, 167*, 25–37.

Beckers, S., & Vennekens, J. (2018). A principled approach to defining actual causation. *Synthese, 195*(2), 835–862.

Blanchard, T., & Schaffer, J. (2017). Cause without default. *Making a Difference*, 175–214.

[15] The cases may but needn't be each other's isomorphs.

[16] However, Samland and Waldmann (2015, 2016) propose that the influence of normality arises from people interpreting the question of causation with the one of normative responsibility. I am partially partial to this proposal.

Clarke, C. (2023). Why your causal intuitions are corrupt: Intermediate and enabling variables. *Erkenntnis*, 1–29.
Clarke, R., Shepherd, J., Stigall, J., Waller, R. R., & Zarpentine, C. (2015). Causation, norms, and omissions: A study of causal judgments. *Philosophical Psychology, 28*(2), 279–293.
Fenton-Glynn, L. (2017). A proposed probabilistic extension of the halpern and pearl definition of 'actual cause'. *The British Journal for the Philosophy of Science*.
Gallow, J. D. (2021). A model-invariant theory of causation. *Philosophical Review, 130*(1), 45–96.
Gerstenberg, T., & Icard, T. (2020). Expectations affect physical causation judgments. *Journal of Experimental Psychology: General, 149*(3), 599.
Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., Spirtes, P., Teng, C. M., & Zhang, J. (2010). Actual causation: a stone soup essay. *Synthese, 175*, 169–192.
Glymour, C., & Wimberly, F. (2007). Actual causes and thought experiments. *Causation and Explanation, 4*, 43.
Hall, N. (2007). Structural equations and causation. *Philosophical Studies, 132*, 109–136.
Halpern, J. Y. (2016). *Actual Causality*. MIT Press.
Halpern, J. Y., & Hitchcock, C. (2010). Actual causation and the art of modeling. In *Probability causality and heuristics: A tribute to Judea Pearl* (pp. 383–406). College Publications.
Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science*.
Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part II: Explanations. *The British Journal for the Philosophy of Science*.
Harinen, T. (2017). Normal causes for normal effects: Reinvigorating the correspondence hypothesis about judgments of actual causation. *Erkenntnis, 82*(6), 1299–1320.
Henne, P., O'Neill, K., Bello, P., Khemlani, S., & De Brigard, F. (2021). Norms affect prospective causal judgments. *Cognitive Science, 45*(1), e12931.
Henne, P., Pinillos, Á., & De Brigard, F. (2017). Cause by omission and norm: Not watering plants. *Australasian Journal of Philosophy, 95*(2), 270–283.
Hiddleston, E. (2005). A causal theory of counterfactuals. *Noûs, 39*(4), 632–657.
Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy, 98*(6), 273–299.
Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *Philosophical Review, 116*(4), 495–532.
Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy, 106*(11), 587–612.
Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. *Moral Psychology, 2*, 441–447.
Kominsky, J. F., & Phillips, J. (2019). Immoral professors and malfunctioning tools: Counterfactual relevance accounts explain the effect of norm violations on causal selection. *Cognitive Science, 43*(11), e12792.
Maslen, C. (2004). Causes, contrasts, and the nontransitivity of causation. In N. Hall, L. A. Paul, & J. Collins (Eds.), *Causation and Counterfactuals* (pp. 341–357). MIT Press.
Maudlin, T. (2004). Causation, counterfactuals, and the third factor. In J. Collins, E. J. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals*. MIT Press.
McDonald, J. (2023). Essential structure for apt causal models. *Australasian Journal of Philosophy* (forthcoming).
McGrath, S. (2005). Causation by omission: A dilemma. *Philosophical Studies, 123*(1–2), 125–148.
Menzies, P. (2004). Causal models, token causation, and processes. *Philosophy of Science, 71*(5), 820–832.
Menzies, P. (2017). The problem of counterfactual isomorphs. In *Making a difference: Essays on the philosophy of causation* (pp. 153–174).
Northcott, R. (2008). Causation and contrast classes. *Philosophical Studies, 139*, 111–123.
Paul, L. A., Hall, N., & Hall, E. J. (2013). *Causation: A user's guide*. Oxford University Press.
Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge University Press.
Rosenberg, I., & Glymour, C. (2018). *Review of joseph halpern, actual causality*.
Ross, L. N., & Woodward, J. F. (2021). Irreversible (one-hit) and reversible (sustaining) causation. *Philosophy of Science*, 1–10.
Samland, J., & Waldmann, M. (2015). Highlighting the causal meaning of causal test questions in contexts of norm violations. In *Proceedings of the 37th annual conference of the cognitive science society*, pp. 2092–2097.

Samland, J., & Waldmann, M. (2016). How prescriptive norms influence causal inferences. *Cognition, 156*, 164–176.
Schaffer, J. (2005). Contrastive causation. *Philosophical Review, 114*(3), 327–358.
Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, Prediction, and Search*. Mit Press.
Thomson, J. J. (2003). Causation: Omissions. *Philosophy and Phenomenological Research, 66*(1), 81–103.
Weslake, B. (2015). A partial theory of actual causation. *British Journal for the Philosophy of Science*.
Wolff, J. E. (2016). Using defaults to understand token causation. *Journal of Philosophy, 113*(1), 5–26.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology*. Oxford University Press.
Wysocki, T. (2020). Normality: A two-faced concept. *Review of Philosophy and Psychology, 11*(4), 689–716.