Why Your Causal Intuitions are Corrupt: Intermediate and Enabling Variables

Christopher Clarke

Forthcoming in *Erkenntnis*

Abstract

When evaluating theories of causation, intuitions should not play a decisive role, not even intuitions in flawlessly-designed thought experiments. Indeed, no coherent theory of causation can respect the typical person's intuitions in redundancy (pre-emption) thought experiments, without disrespecting their intuitions in threat-and-saviour (switching / short-circuit) thought experiments. I provide a deductively sound argument for these claims. Amazingly, this argument assumes absolutely nothing about the nature of causation. I also provide a second argument, whose conclusion is even stronger: the typical person's causal intuitions are thoroughly unreliable. This argument proceeds by raising the neglected question: in what respects is information about intermediate and enabling variables relevant to reliable causal judgment?

1 Intuitions about Causation

This paper is about single-case event causation. This is also known as actual causation—the causal relation that holds between particular events, as opposed to the causal contribution that a particular variable makes to another variable. When philosophers and psychologists elicit peoples' intuitions about single-case event causation, they do so by presenting people with thought experiments. Of course, some thought experiments are misleadingly presented, and some are difficult for the participants to understand, and some end up eliciting hesitant or conflicting responses from the participants. And when a thought experiment is flawed in any of these ways, then the responses elicited by that thought experiment can be ignored, everyone agrees (Collins, Hall, and Paul 2004, sec. 5). So long as a thought experiment is not flawed in any of these ways, however, philosophers have traditionally rejected any theory of causation that conflicts with the intuitions elicited by that thought experiment:

> Work on philosophy of causation is, not surprisingly, heavily driven by intuitions about cases. Standard procedure often seems to be the following: A philosopher proposes a new analysis of causation, showing how it delivers the intuitively correct results about a wide range of cases. But then novel cases are proposed, and intuitions about them exhibited that run counter to the given theory—at which point, either refinements are added to accommodate the recalcitrant "data", or it's back to the drawing board (Collins, Hall, and Paul 2004, 30).
>
> When common sense delivers a firm and uncontroversial answer about a not-too-far-fetched case, theory had better agree (Lewis 1983, 194).

As an illustration of this point, consider the following thought experiment:

> **Redundant Alarm**. The cat leaps on Hanna, and Hanna wakes one second before seven o'clock. If the cat hadn't lept, however, the alarm clock still would have rung at seven, and its ringing would have caused Hanna to wake.

When presented with Redundant Alarm, the typical person is confident that the cat's leaping was a cause of Hanna's waking. And this is despite the fact that Hanna's waking did not counterfactually depend upon the cat's leaping. That is to say, if the cat's leaping had not occurred, then Hanna's waking still would have occurred anyway. The typical person's intuitions in cases such as this one are widely considered to provide a decisive reason to reject the theory that causation is just counterfactual dependence. [1] More recently, however, a few philosophers of causation have sounded a note of caution. In particular, these philosophers have expressed strong sympathies with the following idea:[2]

> **Defeasible**. When evaluating theories of causation, the typical person's intuitions should not count as decisive evidence, not even their intuitions in flawless thought experiments. At best, these intuitions count as defeasible evidence.

Tentative support for Defeasible comes from the concerns that philosophers have voiced about intuitions in general: is there enough homogeneity in people's intuitions to even talk of the typical person's intuitions in the first place? and why think that intuitions are in general reliable? (Alexander and Weinberg 2008; Machery and O'Neill 2014; Machery 2017). Tentative support for Defeasible also comes from the more specific worry that peoples' causal intuitions might be the result of various well-known cognitive biases (Rose 2017). And further inductive support for Defeasible comes from the fact that philosophers, it seems, have repeatedly failed to find a theory of causation that respects the typical person's intuitions. In particular, philosophers have failed to find a single theory of causation that respects the typical person's intuitions in cases like Redundant Alarm, while also respecting the typical person's intuitions in the following sort of case:[3]

> **Boulder Threat and Saviour**. A large boulder is dislodged, and rolls toward Hiker. Before the boulder reaches Hiker, she sees the boulder and ducks. The boulder sails a few centimetres over her head. Hiker survives and then walks home.

In this case, the dislodging of the boulder causes an event (the boulder's traveling towards Hiker) that threatens to injure Hiker, and thus to prevent hiker from walking home. But the dislodging of the boulder also causes a second event (the Hiker's ducking) that saves Hiker from this threat. The typical person judges that the dislodging of the boulder was not a cause of Hiker's walking home. I will call such cases *threat-and-saviour* cases, where threat-and-saviour cases include cases of short-circuits and switching (to use the philosophical jargon).[4] I will contrast threat-and-saviour cases with cases such as Redundant Alarm, which I will call *redundancy* cases, which include cases of pre-emption and overdetermination (to use the jargon again).[5]

To repeat, it seems that philosophers have failed to find a single theory of causation that respects the typical person's intuitions in threat-and-saviour cases, while also respecting their intuitions in redundancy cases. And some philosophers take this as a reason to be sympathetic towards the following idea (Hall 2004):

> **Disunity**. No good theory of causation can respect the typical person's causal intuitions in redundancy thought experiments, without disrespecting their causal intuitions in threat-and-saviour thought experiments.

Nevertheless, these reasons in favour of Defeasible and Disunity are tentative and inductive (see Section 4 for further discussion). For example, the theory offered in Gallow (2021) appears to respect the typical person's intuitions in many of the key redundancy and threat-and-saviour thought experiments. This weakens the case for Disunity.

My aim in the present paper is to provide a compelling argument in favour of Defeasible and Disunity. To do this, I will proceed as follows. Section 2 presents two thought experiments, which I claim are free of flaws. Indeed, the first is a simple threat-and-saviour thought experiment, and the second is a simple redundancy thought experiment. Section 2 also presents the (unsurprising) results of an informal survey of peoples' intuitions about these thought experiments—where these people include both philosophers and non-philosophers. In light of this informal survey, Sections 3–4 provide a deductively sound argument for Disunity. Amazingly, this argument makes no controversial philosophical assumptions. In particular, it assumes absolutely nothing about the nature of causation. The argument proceeds by showing that—insofar as a philosophical theory of causation respects the typical person's intuitions in my two thought experiments—any such theory of causation will be forced to revise its initial causal judgment in at least one of these thought experiments, when extra information about "intermediate" and "enabling" variables is provided.

Section 5 turns from metaphysics to epistemology. Firstly, I will show how Defeasible follows as a corollary of the argument in Section 4. Secondly, I will extend the argument of Section 4, to argue that the typical person's causal intuitions are not only defeasible, but also thoroughly unreliable. This extended argument makes one assumption about the nature of causation—an assumption which will be somewhat controversial, I expect.

However, this argument avoids relying on extremely controversial assumptions about causation, such as the assumption that facts about causation are determined by facts about counterfactual dependence alone, or the assumption that absence causation is possible. Section 6 explores the extent to which one can resist the argument in Section 5, if one insists that causal knowledge is norm-involving. I suggest that the argument in Section 5 is not easy to resist. At any rate, even if this argument can be resisted, it at very least demonstrates that, to shore up their position, defenders of the reliability of causal intuitions need to answer the following question: in what respects is information about enabling and intermediate variables relevant to reliable causal judgment? and why?

2 Two Thought Experiments Presented

Jo's Thought Experiment

This thought experiment involves three neuro-chemicals: a stimulant, an antidote to this stimulant, and a chemical called disruptase. In the case of a person called Jo, the three corresponding variables (the stimulant level in her brain, the antidote level in her brain, and the disruptase level in her brain) enter into the following counterfactual dependence relationships:

(1) Jo will be awake at midnight if and only if disruptase is present (shortly before midnight). Otherwise, she will fall asleep at midnight.

(2) Disruptase is present (shortly before midnight) if and only if the stimulant is present and the antidote to this stimulant is absent.

(3) Zeus mints contain this stimulant, and indeed this stimulant is present if and only if Jo has recently eaten a Zeus mint.

(4) Zeus mints also contain the antidote, and indeed this antidote is present if and only if Jo has recently eaten a Zeus mint.

Pedantic warning: in describing my thought experiments, these "if and only if" statements are not used to express mere material biconditionals. Instead, they are used to express counterfactual dependence relationships. For example, I stipulate that the proper way to read statement 1 is this: "Jo will be awake at midnight if and only if disruptase is present" is true of Jo for any day of the year, and it would remain true of Jo, even under hypothetical interventions to any of the variables explicitly mentioned in statements 1, 2, 3, and 4. [6]

Let's now imagine a particular evening in Jo's life:

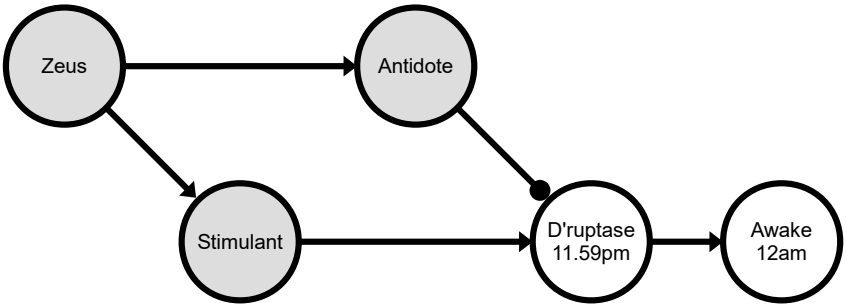

Figure 1: Mint Threat

> At the start of the evening, the following are all absent: disruptase, the stimulant, and the antidote. Jo then eats a Zeus mint. Both the stimulant and the antidote are then present. Disruptase remains absent, even as midnight arrives. Jo falls asleep at midnight.

In sum, this thought experiment provides some particular information about what events occurred that evening, and also some variable-level counterfactual information about Jo, namely counterfactual dependence relationships 1–4, which involve these three neurochemical variables in her brain. Given this information, what does the typical person say about what caused what that particular evening? I've found that the typical person strongly disagrees with the claim that "eating a Zeus mint was a cause of Jo's falling asleep that midnight". [7] This is true for both philosophers and non-philosophers. This is unsurprising, because this thought experiment is clearly a threat-and-saviour thought experiment. For this reason, I will label this thought experiment Mint Threat.

For ease of reference, I've represented the counterfactual dependence relationships in Mint Threat in fig. 1. The lines in this diagram contain either a normal arrow or a circular arrow. Interpret the arrows in this figure similarly to the arrows in so-called neuron diagrams. That is to say, for any given variable V in this figure, variable V will be "on" if and only if (i) there is at least one normal arrow pointing to V from a variable that is "on", and (ii) there is no circular arrow pointing to V from any variable that is "on". In this figure, the variables that actually took the value "on" are shaded, and the variables that actually took the value "off" are unshaded. This figure was not presented to participants in the experiment.

Anna's Thought Experiment

Let's now imagine a different thought experiment. This thought experiment involves two neuro-chemicals: hypnotase *A* and hypnotase *B*. In the case of a person called Anna, the two corresponding variables (the hypnotase *A* level in her brain and the hypnotase *B* level in her brain) enter into the following counterfactual dependence relationships:

(5) Anna will fall asleep at midnight if and only either hypnotase *A* or hypnotase *B* are present (shortly before midnight). Otherwise, she will be awake at midnight.

(6) Hypnotase *A* is present if and only if Anna has recently eaten an amber pill.

(7) Hypnotase *B* is present if and only if Anna has recently eaten a black pill and Anna has not recently eaten an amber pill.

Let's now imagine a particular evening in Anna's life:

> At the start of the evening, hypnotase *A* and *B* are both absent. Anna then eats an amber pill and a black pill. Hypnotase *A* is then present, but hypnotase *B* stays absent. Anna falls asleep at midnight.

Given this information, what does the typical person say about what caused what that particular evening? I've found that the typical person strongly agrees with the claim that "eating the amber pill was a cause of Anna's falling asleep that midnight", [8] although they disagree with the claim that "eating the black pill was a cause of Anna's falling asleep at midnight". [9] This is clearly a redundancy thought experiment, and as such I will label it Redundant Pill.

For ease of reference, I've represented the counterfactual dependence relationships in Redundant Pill in fig. 2. Again, this figure was not presented to participants in the experiment. (The arrow pointing into the "awake 12am" node should be thought of as coming from some unspecified variable that is always "on". I needed to add this arrow to this diagram to ensure that the diagram says that Anna will be awake at midnight, unless either hypnotase *A* or hypnotase *B* is present.)

3 A Hybrid Thought Experiment

To make my case for Defeasible and Disunity, this section will present a thought experiment that I will call Hybrid. In this section, I will show that Hybrid is the thought experiment that arises when one takes Mint Threat and then adds some extra information about how the counterfactual dependencies in Mint Threat are "mediated" and "enabled" by other variables. Similarly, I will show that Hybrid is the thought experiment

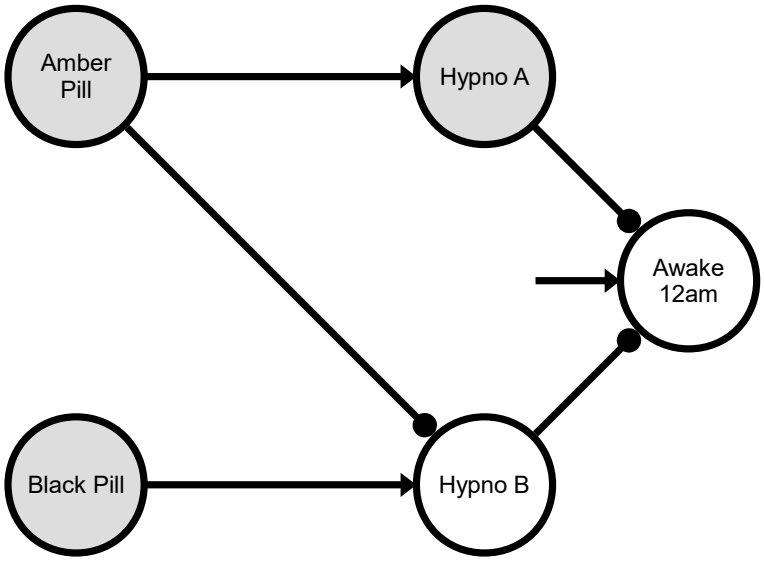

Figure 2: Redundant Pill

that arises when one takes Redundant Pill and then adds some extra information about how the counterfactual dependencies in Redundant Pill are mediated by other variables. In this respect, Mint Threat and Redundant Pill are both stripped down versions of Hybrid, so to speak. Given this relationship between Mint Threat and Redundant Pill, Section 4 will establish Disunity and Defeasible. Section 5 will then present a similar argument, whose conclusion is that the typical person's causal intuitions are thoroughly unreliable.

The Hybrid thought experiment involves four chemicals: a stimulant, hypnotase *A*, hypnotase *B* and disruptase. Hybrid says that, in the case of someone called Joanna, the four corresponding variables enter into the following counterfactual dependence relationships. These details are complex, so I will provide a more intuitive gloss on these complex details in just a moment. Refer also to fig. 3.

(I) Joanna will be awake at midnight if and only if disruptase is present (shortly before midnight). Otherwise, she will fall asleep at midnight.

(II) Disruptase is present (shortly before midnight) if and only both hypnotase *A* and hypnotase *B* are absent.

(III) Hypnotase *A* is present if and only if Joanna has recently eaten an amber pill.

(IV) Hypnotase *B* is present if and only if Joanna has recently eaten a black pill and the stimulant is absent.

(V) The stimulant is present if and only if Joanna has recently eaten an amber pill.

(VI) Joanna is also called Jo and is also called Anna. Zeus mints are also called amber pills. Hypnotase *A* is also called the antidote.

What's more, Hybrid says that on the particular evening in question:

> At the beginning of the evening, the following are all absent: the stimulant, hypnotase *A*, and hypnotase *B*. Joanna then eats an amber pill and a black pill. Hypnotase *A* and the stimulant are then present. But hypnotase *B* stays absent. Just before midnight, disruptase is absent. Anna falls asleep at midnight.

As you can see from the above, Hybrid is complicated. So let me give a bit more of an intuitive sense of what is going on in Hybrid. (Refer also to fig. 3.) Eating the black pill is necessary for hypnotase *B* to be present (says *IV*) and hypnotase *B* being present is in turn sufficient for Joanna to fall asleep at midnight (says *I* and *II*). In this sense, the black pill can act as a sedative. But eating the amber pill is sufficient for hypnotase *B* to be absent (says *IV* and *V*). In this sense, the amber pill blocks the sedative power of the black pill, and thus can act as a stimulant. But the amber pill is also sufficient for hypnotase *A* to be present (says *III*) which in turn is sufficient for Joanna to fall asleep

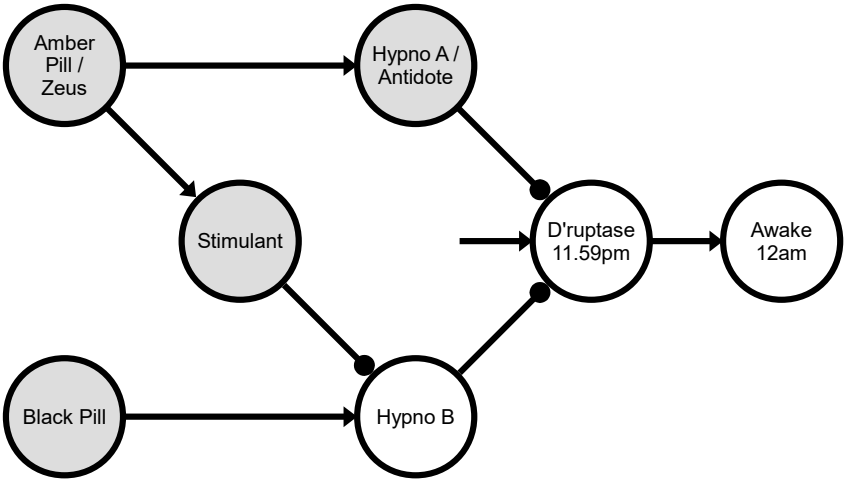

Figure 3: Hybrid

at midnight (says *I* and *II*). In this sense, the amber pill is so powerful a sedative that it acts as an antidote to its own blocking of the sedative power of the black pill.

Of course, the gloss I've just given on Hybrid is very rough. So it's important to point out that the argument of the present paper will rely only on *I–VI* themselves, not on the rough gloss that I've just given on them. What's more, my argument will only rely on peoples' intuitions about Mint Threat and Redundant Pill. Peoples' intuitions about Hybrid don't matter. Instead, the only role that Hybrid plays in my argument is to show that Mint Threat and Redundant Pill bear some similarity to each other, to some degree.

In what respect does Hybrid do this? On inspection, it is obvious that Hybrid (which is defined by information *I–VI*) contains all the information that defines Mint Threat (namely *1–4*). It's also obvious that Hybrid contains some extra information over and above Mint Threat. It's a straightforward task to confirm that this extra information is namely:

(a) There is at least one variable I_1 that "mediates" the counterfactual dependence of the disruptase variable upon the stimulant variable. By this I mean:

  (i) Disruptase is present (shortly before midnight) if and only if I_1 is absent and the antidote is absent.

  (ii) I_1 is absent if and only if the stimulant is present.

(b) There is at least one set of facts F that obtain on the evening in question and that "enabled" the role played by the stimulant variable. By this I mean:

   (iii) Had each fact in set F instead not obtained, *ii* above would be false. Specifically, I_1 would have been absent, regardless of the stimulant level; but all the other counterfactual dependencies (namely the arrows in fig. 3) would have remained the same.

(c) One set of facts that satisfies the description F above is the singleton set {Joanna has recently eaten a black pill}.

(d) One variable that satisfies the description I_1 above is called hypnotase B.

(e) The antidote is also called hypnotase A.

(f) Joanna is also called Jo and is also called Anna. Zeus mints are also called amber pills.

To see this, note: when one adds information *a* and *b* to fig. 1, one gets fig. 4; when one then adds information *c*–*f* in addition, one gets fig. 3. It is worth emphasising that this information *a*–*f* is not information about causation; instead it is information about counterfactual dependence. So Hybrid itself makes no assumptions about causation. This is important because it shows that one cannot object to my arguments in Sections 4 and 5 by claiming that Hybrid itself embodies any controversial assumptions about causation. It does not.

Similarly, on inspection, it is obvious that Hybrid contains all the information that defines Redundant Pill (namely *5*–*7*). It's also obvious that Hybrid also contains some extra information over and above Redundant Pill. Indeed, it's a straightforward task to confirm that this extra information is namely:

(g) There is at least one variable I_2 that "mediates" the counterfactual dependence of the sleep variable upon the hypnotase A and hypnotase B variables. By this I mean:

   (i) Joanna will fall asleep at midnight if and only if I_2 is absent (shortly before midnight). Otherwise, Joanna will be awake at midnight.

   (ii) I_2 is absent (shortly before midnight) if and only if either hypnotase A or hypnotase B is present.

(h) There is at least one variable I_3 that "mediates" the counterfactual dependence of the hypnotase B variable upon the amber pill variable. By this I mean:

   (i) Hypnotase B is present if and only if Joanna has recently eaten the black pill and I_3 is absent.

   (ii) I_3 is absent if and only if Joanna has not recently eaten the amber pill.

(i) One variable that satisfies the description I_2 above is called disruptase.

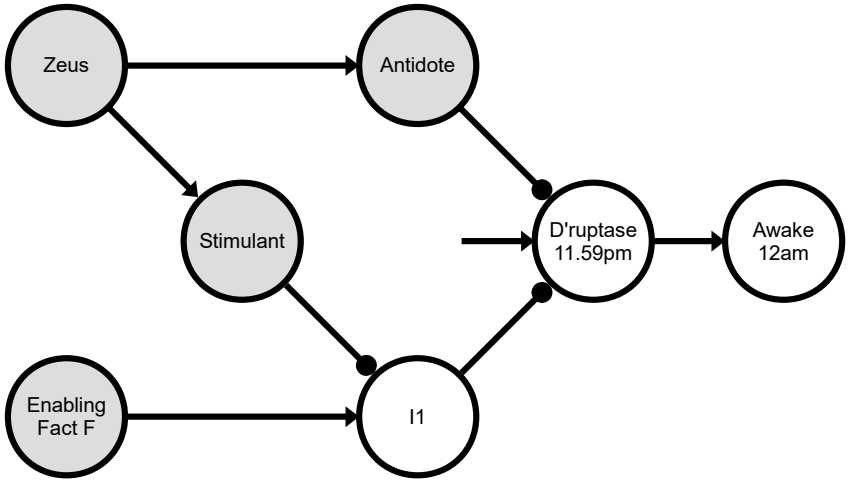

Figure 4: Mint Threat plus *a* and *b*

(j) One variable that satisfies the description I_3 above is called the stimulant.

(k) Joanna is also called Jo and is also called Anna. Zeus mints are also called amber pills.

To see this, note: when one adds information *g* and *h* to fig. 2, one gets fig. 5; when one then adds information *i*–*k* in addition, one gets fig. 3. Again, it is worth noting that this information *g*–*k* is not information about causation; instead it is information about counterfactual dependence. To repeat, Hybrid itself makes no assumptions about causation.

4 The Disunity of the Metaphysics of Causation

In this section, I will use the existence of Hybrid to establish Disunity: no coherent metaphysical theory of causation can both (a) issue a decisive causal verdict about Mint Threat in line with the typical person's intuitions, and (b) issue a decisive causal verdict about Redundant Pill in line with the typical person's intuitions.

Here's the quick outline of my argument. Metaphysical theories of causation describe what causation is. As such, so long as one describes a given case in sufficient detail, a metaphysical theory will issue a decisive verdict about whether *C* was a cause of *E* in that case. It follows that, whenever a metaphysical theory of causation issues a

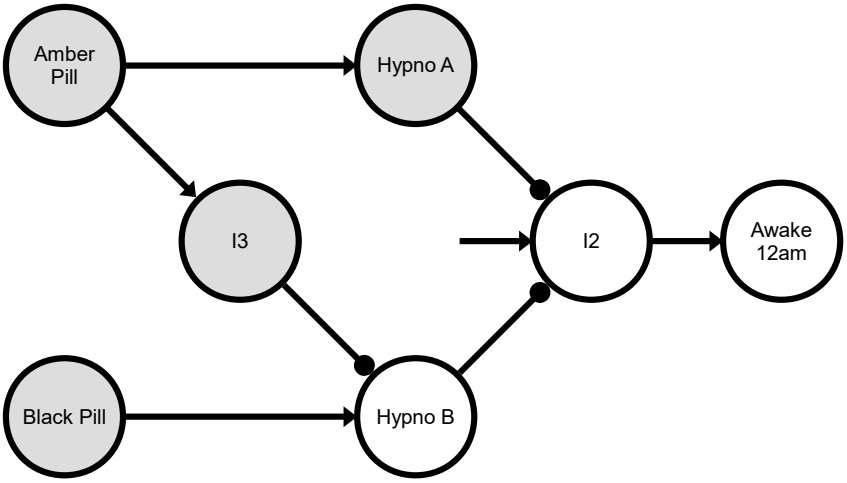

Figure 5: Redundant Pill plus *g* and *h*

decisive verdict about a particular case, this metaphysical theory should not then reverse this decisive verdict, when one adds more information about the enabling and intervening variables in this case. This no-reversals condition is a condition that any theory of causation must meet, if it is to be a coherent metaphysical theory. See Hitchcock (2009, 398), Paul and Hall (2013, 161–62), Halpern (2016, chap. 4) and Gallow (2021) for agreement, and for discussion of this point.

But the existence of Hybrid generates a dilemma for metaphysical theories of causation (insofar as they do justice to the typical person's intuitions in both Redundant Pill and Mint Threat). On the one hand, a metaphysical theory of causation could issue the same causal verdict in Hybrid as it does in Mint Threat—in which case its verdict in Hybrid would be the reverse of its verdict in Redundant Pill. On the other hand, a metaphysical theory of causation could issue the same causal verdict in Hybrid as it does in Redundant Pill—in which case its verdict in Hybrid would be the reverse of its verdict in Mint Threat. So either one's theory will reverse its verdict about Mint Threat (when one adds the extra information contained in Hybrid) or it will reverse its verdict about Redundant Pill (when one adds the extra information contained in Hybrid). Therefore, no metaphysical theory of causation can possibly meet the no-reversals condition (insofar as it does justice to the typical person's intuitions in both Redundant Pill and Mint Threat).

I will now present this simple argument in a more rigorous form. To do so, I should

first explain what I mean when I say that a metaphysical theory of causation *T* issues a decisive verdict (for a case described by information *I*) on whether event *C* was a cause of event *E* in that case. By this I mean: either (a) theory *T* plus information *I* deductively entails that event *C* was a cause of event *E*, or (b) theory *T* plus information *I* deductively entails that event *C* was not a cause of event *E*. With this definition of "issuing a decisive verdict" in mind, consider a metaphysical theory of causation *T* that issues a decisive causal verdict about Mint Threat in line with the typical person's intuitions, and that also issues a decisive causal verdict about Redundant Pill in line with the typical person's intuitions. My argument proceeds as follows:

1. The typical person's intuitions in Mint Threat are that the Zeus mint was not a cause of Jo falling asleep. (Assumption supported by my empirical survey.)

2. So theory *T*, combined with the information that defines Mint Threat, deductively entails that the Zeus mint was not a cause of Jo falling asleep. (From 1, and the definition of theory *T*.)

3. So theory *T*, combined with Mint Threat plus the information *a–f*, deductively entails that the Zeus mint was not a cause of Jo falling asleep. (From 2, and the nature of deductive entailment.)

4. So theory *T*, combined with Mint Threat plus the information *a–f*, deductively entails that the Zeus mint / amber pill was not a cause of Joanna falling asleep. (From 3, and the definition of information *f*, which says that Jo = Joanna, and that Zeus mint = amber pill.)

5. But the information that defines the Hybrid thought experiment is equivalent to Mint Threat plus information *a-f*. (Noted in the last section.)

6. So theory *T*, combined with the information that defines Hybrid, deductively entails that the Zeus mint / amber pill was not a cause of Joanna falling asleep. (From 4 and 5.)

We can do something similar for Redundant Pill:

7. The typical person's intuitions in Redundant Pill are that the amber pill was a cause of Anna falling asleep. (Assumption supported by my empirical survey.)

8. So theory *T*, combined with the information that defines Redundant Pill, deductively entails that the amber pill was a cause of Anna falling asleep. (From 7, plus the definition of theory *T*.)

9. So theory *T*, combined with Redundant Pill plus the information *g–k*, deductively entails that the amber pill was a cause of Anna falling asleep. (From 8, and the nature of deductive entailment.)

10. So theory *T*, combined with Redundant Pill plus the information *g–k*, deductively entails that the Zeus mint / amber pill was a cause of Joanna falling asleep. (From 9, and the definition of information *k*, which says that Anna = Joanna, and that Zeus mint = amber pill.)

11. But the information that defines the Hybrid thought experiment is equivalent to Redundant Pill plus information *g-k*. (Noted in the last section.)

12. So theory *T*, combined with the information that defines Hybrid, deductively entails that the Zeus mint / amber pill was a cause of Joanna falling asleep. (From 10 and 11.)

Putting this altogether:

13. Theory *T*, combined with the information that defines Hybrid, deductively entails a contradiction: the Zeus mint / amber pill both was and was not a cause of Joanna falling asleep. (From 6 and 12.)

14. But Hybrid describes a possible situation. (Assumption.)

15. Whenever a metaphysical theory deductively entails a contradiction, when applied to a possible situation, then the metaphysical theory is incoherent. (Assumption)

16. So metaphysical theory *T* is incoherent. (From 13, 14 and 15.)

17. So Disunity is true: no coherent metaphysical theory of causation can both (a) issue a decisive causal verdict about Mint Threat in line with the typical person's intuitions, and (b) issue a decisive causal verdict about Redundant Pill in line with the typical person's intuitions. (From 16, and the definition of *T*.)

Implications

How can this argument for Disunity be resisted? Note that this argument makes no assumptions about the typical person's intuitions regarding the Hybrid thought experiment. So this argument for Disunity cannot be rejected by rejecting its claims about peoples' intuitions about Hybrid. This argument makes no such claims. Instead, this argument for Disunity relies only on the following assumptions: two empirical assumptions about the typical person's intuitions (premises 1 and 7), the assumption that Hybrid describes a possible situation (premise 14), and a principle about incoherence (premise 15). But the latter principle about incoherence is uncontroversial. And it is undeniable that Hybrid describes a possible situation. (As I noted in Section 3, Hybrid describes various counterfactual dependencies. Hybrid does not embody any controversial assumptions about causation. In fact, it doesn't embody any assumptions about causation at all. So it is undeniable that Hybrid describes a possible situation.) Therefore, I cannot see any reasonable way to deny the assumptions of my argument—other than by

challenging the empirical results of my survey, namely by performing a more rigorous survey of peoples' intuitions.

This is a considerable advance on the existing arguments for Disunity in the literature. Firstly, several philosophers have already pointed out that there is an analogy between redundancy cases such as Redundant Alarm and threat-and-saviour cases such as Boulder Threat-and-Saviour (Hall 2007; Weslake, n.d., sec. 3). The events that occur in each case enter into perfectly analogous counterfactual dependence relationships with each other. [10] Despite this, the typical person does not judge the causes in each case analogously: (C) the pushing of the boulder is not a cause of (E) hiker walking home; but (C') Hanna's cat's leaping is a cause of (E') Hanna's waking. Menzies (2017) calls this phenomenon the problem of counterfactual isomorphs. Similarly, Halpern and Hitchcock (2015) call this phenomenon the problem of isomorphism. This phenomenon entails that either (i) the typical person's intuitions in some redundancy cases are false, or (ii) the typical person's intuitions in some threat-and-saviour cases are false, or (iii) facts about single-case event causation (between particular events) are not determined by facts about counterfactual dependence (between variables) alone. Since *i* and *ii* entail Disunity, this lends some support to Disunity.

Nevertheless, one can still defend Disunity, in the face of this phenomenon, if one insists that (iii) facts about causation are not determined by facts about counterfactual dependence alone.[11] In contrast, note that my argument above makes absolutely no assumptions about the nature of causation. In particular, my argument did not assume, for example, that facts about causation are determined solely by facts about counterfactual dependence. So my argument offers a much more powerful argument for Disunity.

Secondly, consider Hall's (2004) hypothesis that our concept of causation is actually two incompatible concepts in disguise. On the one hand, there's the conception of causation as counterfactual dependence. An event E is caused by an event C if and only if: if C had been absent, then E would have been absent too. On the other hand, there's the conception of causation as some sort of intrinsic and transitive relation between events.[12] Hall notes that the fomer conception of causation (as counterfactual dependence) is incompatible with the latter conception of causation (as an intrinsic and transitive relation).[13] Thus we have two incompatible conceptions of causation on the table. Now, the conception of causation as counterfactual dependence is compatible with the typical person's intuitions in most threat-and-saviour cases, but not in most redundancy cases, Hall notes. But the conception of causation as some sort of transitive and intrinsic relation is compatible with the typical persons' intuitions in most redundancy cases, but not with their intuitions in most threat-and-saviour cases, Hall notes. Neither conception can do justice to the typical person's intuitions in both types of cases. But let's suppose for one moment that these two conceptions of causation exhaust all the good conceptions of causation. Disunity follows: there is no good conception of causation that is compatible with the typical person's intuitions in both redundancy and

threat-and-saviour cases.

The obvious way of objecting to Hall's argument above is to question his supposition that that these two conceptions of causation exhaust all the good conceptions of causation. For all that Hall says, there may be an alternative conception of causation that can do justice to all the causal intuitions of the typical person; it's just that philosophers haven't found this alternative conception of causation yet. Therefore, as Paul and Hall (2013, 248) readily acknowledge, Disunity remains currently an open question in the literature.

However, my argument above shows how Disunity can be established without the need for Hall's supposition. My argument shows that no coherent theory of causation can respect the typical person's causal intuitions in redundancy thought experiments, without disrespecting her intuitions in threat-and-saviour thought experiments.

In sum, my argument advances our understanding of causation by providing a compelling argument for Disunity—an argument that makes no controversial philosophical assumptions. It follows that the typical person's intuitions in at least some threat-and-saviour cases (Mint Threat for one) are simply not compatible with their intuitions in at least some redundancy cases (Redundant Pill for one). Either one's metaphysical theory of causation will end up disrespecting peoples' intuitions in Mint Threat (and probably by extension other threat-and-saviour cases such as Boulder Threat). Or one's theory will end up disrespecting peoples' intuitions in Redundant Pill (and probably by extension other redundancy cases such as Redundant Alarm). Philosophers of causation have been aiming to do the impossible: they have been aiming to unite our intuitions about both types of cases within a single metaphysical theory.

5 The Epistemology of Causal Judgement

I hope that Section 4 has convinced you of Disunity. Whereas the main focus of Section 4 was on metaphysical theories of causation, the focus of Section 5 will instead be the epistemology of causal judgement. In Section 5 I will do two things. Firstly, I will show that Defeasible follows as a corollary of the argument in Section 4. Defeasible says that not all of the typical person's causal intuitions (in flawless thought experiments) are indefeasible evidence, when it comes to evaluating claims about causation. Secondly, I will argue for a bolder claim:

> **Thoroughly Unreliable**. The typical person's causal intuitions are not just defeasible; they are thoroughly unreliable.

An Argument for Defeasible

1. Indefeasible evidence means evidence that continues to warrant the same causal judgment, even when one is provided with more information about the case under examination. (Definition.)

2. The typical person's intuitions in (a) Mint Threat and (b) Redundant Pill are indefeasible evidence. (Suppose for reductio ad absurdum.)

3. So if the typical person were presented with the extra information in the Hybrid thought experiment (about intermediate and enabling variables), then the typical person would continue to be warranted in judging that (a) the Zeus mint was not a cause of Jo falling asleep, and (b) the amber pill was a cause of Anna falling asleep. (From 1 and 2 and the argument for Disunity.)

4. But *a* and *b* are logically incompatible, given the extra information in Hybrid, that Jo = Anna = Joanna, and that Zeus mints are amber pills. (From the definition of Hybrid.)

5. So, when presented with the extra information Hybrid, one is not warranted in accepting both *a* and *b*. (From 4.)

6. So the typical person's intuitions in at least one of (a) Mint Threat and (b) Redundant Pill are (at best) defeasible evidence. (From reductio ad absurdum of our supposition 2, which contradicts 5.)

The Unreliability of Causal Intuition

It's now time to make the case that the typical person's causal intuitions are thoroughly unreliable. To make this case, my strategy will be as follows. I will argue that the extra information that Hybrid contains over and above Mint Threat is negligible: this extra information will not result in a reliable inquirer making different causal judgments about Hybrid than the causal judgments she makes about Mint Threat. Similarly, I will argue that the extra information that Hybrid contains over and above Redundant Pill is negligible: this extra information will not result in a reliable inquirer making different causal judgments about Hybrid than the causal judgments she makes about Mint Threat. It follows that a reliable inquirer will make the same causal judgment in Mint Threat as she does in Redundant Pill. But, as Section 2 noted, the typical person has different causal intuitions in these two thought experiments. And so the typical person's causal intuitions are unreliable, I conclude. That's a very quick sketch of my argument for Thoroughly Unreliable.

My argument for Thoroughly Unreliable will be a long one. So it's worth stating up-front the basic four assumptions upon which my argument will rely:

Mediation Assumption (**Rough Version**). Counterfactual dependencies in bio-medicine are often mediated by hundreds of chemicals, and indeed are often mediated by absences.

Enabling Assumption (**Rough Version**). Counterfactual dependencies in bio-medicine are often enabled.

Bayesian Assumption. A reliable inquirer's updated credences (on learning new information) conform to Bayes rule.

Many Causally Equivalent Enablers (**Rough Version**). There are a large list of "putative enablers" that should be treated equivalently to the black pill for the purposes of causal inference.

So, if you object to my argument for Thoroughly Unreliable, you will need to reject at least one of these basic assumptions. It is not enough merely to claim that you think that there is some "disanalogy" between Mint Threat and Hybrid or between Redundant Pill and Hybrid. What's more, note that causation may well differ from counterfactual dependence. Thus my four basic assumptions do not entail any of the following four propositions:

i) Causal relationships in biomedicine are often mediated, and indeed mediated by absences.

ii) Causal relationships can hold between absences.

iii) Causal relationships in biomedicine are often enabled.

iv) Facts about causation are determined by facts about counterfactual dependence alone.

Since my argument for Thoroughly Unreliable will not rely on any of *i–iv*, one cannot object to my argument by rejecting any of *i–iv*. Instead, one has to reject one of my four basic assumptions. Many Causally Equivalent Enablers is the most vulnerable of these basic assumptions. Indeed, it is the only assumption that I will make that is specifically about causation itself. In Section 6 I will consider one way in which one might object to Many Causally Equivalent Enablers.

The Argument in More Detail

My argument for Thoroughly Unreliable begins by examining what happens when one adds extra information to Mint Threat, bit by bit:

1. When presented with Mint Threat, a reliable inquirer will have a low credence in the proposition P_{MINT}, the proposition that Jo's eating a Zeus mint was a cause of her falling asleep that midnight. (Assumption made temporarily for conditional proof.)

2. Adding information *a* to Mint Threat will not increase a reliable inquirer's low credence in P_{MINT}, at least not by very much. (This follows from 1 and the Mediation Assumption and Bayesian Assumption, I will show.)

3. Adding information *b* to Mint Threat (plus information *a*) will not increase her credence, at least not by very much. (This follows from 1 and the Enabling Assumption and Bayesian Assumption, I will show.)

4. Adding information *c* to Mint Threat (plus information *a–b*) will not increase her credence, at least not by very much. (This follows from 1 and More Causally Equivalent Enablers and Bayesian Assumption, I will show.)

5. Adding information *d–f* to Mint Threat (plus information *a–c*) will not change her credence at all. (Assumption.)

6. So, if a reliable inquirer will have a low credence in P_{MINT} when presented with Mint Threat, then a reliable inquirer will have a low-ish credence in P_{MINT} when presented with Mint Threat plus information *a–f*. (From 2 3 4 and 5. This discharges temporary assumption 1.)

7. So, if a reliable inquirer will have a low credence in P_{MINT} when presented with Mint Threat, then a reliable inquirer will have a low-ish credence in P_{MINT} when presented with Hybrid. (From 6, and the fact that Hybrid is equivalent to Mint Threat plus information *a–f*.)

My argument continues by examining what happens when one adds extra information Redundant Pill bit by bit:

8. When presented with Redundant Pill, a reliable inquirer will have a high credence in the proposition P_{AMBER}, the proposition that Anna's eating the amber pill was a cause of her falling asleep that midnight. (Assumption made temporarily for conditional proof.)

9. Adding information *g* and *h* to Redundant Pill will not decrease a reliable inquirer's high credence in P_{AMBER}, at least not by very much. (This follows from 8 and the Mediation Assumption and Bayesian Assumption, I will show.)

10. Adding information *i–k* to Redundant Pill (plus information *g–h*) will not change her credence at all. (Assumption.)

11. So, if a reliable inquirer will have a high credence in P_{AMBER} when presented with Redundant Pill, then a reliable inquirer will have a high-ish credence in P_{AMBER} when presented with Redundant Pill plus information *g–k*. (From 9 and 10. This discharges temporary assumption 8.)

12. So, if a reliable inquirer will have a high credence in P_{AMBER} when presented with Redundant Pill, then a reliable inquirer will have a high-ish credence in P_{AMBER} when presented with Hybrid. (From 11, and the fact that Hybrid is equivalent to Redundant Pill plus information *g–k*.)

13. So, if a reliable inquirer will have a high credence in P_{AMBER} when presented with Redundant Pill, then a reliable inquirer will have a high-ish credence in P_{MINT} when presented with Hybrid. (From 12, and the fact that, according to Hybrid, Zeus mints are amber pills.)

It follows:

14. If a reliable inquirer will *not* have a high-ish credence in P_{MINT} when presented with Hybrid, then a reliable inquirer will *not* have a high credence in P_{AMBER} when presented with Redundant Pill. (From contraposition of 13.)

15. So, if a reliable inquirer will have a low credence in P_{MINT} when presented with Mint Threat, then a reliable inquirer will not have a high credence in P_{AMBER} when presented with Redundant Pill. (From 7 and 14.)

16. The typical person has high credence in P_{AMBER} when presented with Redundant Pill, but low credence in P_{MINT} when presented with Mint Threat. (Assumption supported by the empirical survey in section 2.)

17. The typical person's intuitions in at least one of Mint Threat and Redundant Pill are not those of a reliable inquirer. (From 15 and 16.)

But Mint Threat and Redundant Pill are flawless thought experiments. So the typical person's causal intuitions are thoroughly unreliable, I conclude. That's the overview of my argument. The task that remains, of course, is to use my four basic assumptions to argue for premises 2, 3, 4, 5, 9 and 10. I will now consider each of these premises in turn.

Premise 2: Adding Information *a*

Consider a reliable inquirer who knows the information that defines Mint Threat. How will her causal judgments change when she learns information *a*? Mint Threat already tells her that: (2) disruptase is present if and only if the stimulant is present and the antidote to this stimulant is absent. Information *a* adds to this by telling her that there was at least one variable I_1 that mediates this counterfactual dependence. To be precise, information *a* says that (i) disruptase is present if and only if I_1 is absent and the antidote is absent. And information *a* also says that (ii) I_1 is absent if and only if the stimulant is present.

However, when presented with Mint Threat, a reliable inquirer will already have a high credence that counterfactual dependence 2 is mediated in this way, I will now suggest. After all, whenever there is a counterfactual dependence between any variable at one point in time (the disruptase level) and any variable earlier in time (the stimulant level) then it's likely there will be a variable I at an intermediate point in time

that mediates this dependence. I do not deny that counterfactual dependence between variables is often unmediated through space; for example, this was how gravitation was viewed in the 18th and 19th centuries. But it is rare for counterfactual dependence to be unmediated through time, I assume—rare if not impossible (Ingthorsson 2007). More specifically, in the context of biochemistry, it is typical for the counterfactual dependence between one chemical and a second chemical to be mediated by the presence / absence of a third chemical, I claim. In fact, I claim, it is typical that the counterfactual dependence between one chemical (call it C_0) and a second chemical (call it C_{101}) is mediated by a hundred distinct chemicals: $C_1, C_2, C_3, C_4, \ldots C_{100}$. (The idea is that C_1 mediates the counterfactual dependence between C_0 and C_2; and C_2 mediates the counterfactual dependence between C_1 and C_3; and so on.) This claim is part of what I will call the Mediation Assumption. Therefore, information a amounts to the following: out of the one hundred plus intermediary chemicals between the stimulant variable and the disruptase variable, at least one of the intermediary variables (C_{44} say) is of the type "chemical C_{44} is absent", rather than of the type "chemical C_{44} is present". But it is not at all uncommon in biochemistry for a chemical to be absent if and only some other chemical is present, I'd also claim. This claim is also part of what I call the Mediation Assumption. So it's likely (let's say at least 80 percent likely) that at least one of these hundreds of intermediary variables belongs to the type "chemical C is absent", as information a suggests. To summarize: many counterfactual dependence relationships between two chemicals are mediated by the absence of some third chemical, and this is especially true in the social and biological sciences. [14] Thus a reliable inquirer, when presented with Mint Threat, will already have a high credence that a is true (at least 80 percent let's say).

However, information that is likely is not very consequential. For example, learning that one has *not* won the lottery will not change one's credence in any proposition much at all. More precisely, one can prove, using my Bayesian Assumption, that if a reliable inquirer's credence in P_{MINT} increases when she learns information a, then this increase in credence will be, at most: [15]

$$Cr(P_{MINT})\frac{1 - Cr(A)}{Cr(A)}$$

Here Cr denotes this inquirer's prior credences—her degrees of confidence in various propositions when presented with Mint Threat alone. $Cr(P_{MINT})$ denotes this inquirer's prior credence in P_{MINT} and $Cr(A)$ denotes this inquirer's prior credence that a is true. For example, when $Cr(P_{MINT})$ is 10 percent, and when $Cr(A)$ is 80 percent, then the maximum possible increase is $.10(.20/.80) = .025$, namely 2.5 percent. So a reliable inquirer's credence in P_{MINT}, when presented with Mint Threat plus information a, is at most 12.5 percent in this example. This shows, more generally, that adding information a to Mint Threat will not increase a reliable inquirer's credence in P_{MINT} by very much

at all. This is premise two of my overall argument.

Notice, crucially that my argument for premise two relied only on the following things:

i) The assumption that P_{MINT} is low, 10 percent for example. (This is premise one from my main argument. It is assumed only temporarily for conditional proof.)

ii) The Mediation Assumption about counterfactual dependence, which ensures that information *a* has a high prior credence, 80 percent for example.

iii) The Bayesian Assumption about how a reliable inquirer updates her beliefs.

So my argument for premise two assumes absolutely nothing about causation as such. After all, the Bayesian Assumption is a general epistemological principle. And the Mediation Assumption is an assumption about counterfactual dependence. It is not an assumption about causation as such. The Mediation Assumption says nothing about how causation is mediated, or about whether causation can be mediated by absences. As a result, my argument for premise two holds for any view of causation whatsoever. For example, it hold for views of causation for which causation is not determined by facts about counterfactual dependence, for example. And it holds for views of causation in which absence causation is impossible, to take another example. (It also worth recalling the point I made in Section 3 that information *a*–*k* doesn't say anything about causation as such. It too is just about counterfactual dependence.)

Premise 3: Adding Information *b*

What happens to a reliable inquirer's causal judgments when one adds information *b*? Information *b* says that there is at least one set of facts *F* (that obtained on the evening in question) that "enabled" the role played by the stimulant variable. But, when presented with Mint Threat plus information *a*, a reliable inquirer will already have a high credence that *b* is the case, I will now suggest.

How so? The human body is a complex physiological system. Behind every counterfactual dependence relationship between any two physiological variables, there will be hundreds of sets of biochemical facts, where each set enables this counterfactual dependence on any given occasion. By a set of facts that enable a counterfactual dependence relationship on a particular occasion, I mean: each of the facts in the set actually obtained on that occasion, but if all the facts in this set had instead been absent on that occasion, then the counterfactual dependence in question would not have obtained on that occasion. I call this claim the Enabling Assumption. So given Mint Threat and information *a*, and given the existence of these hundreds of sets of facts, it's likely that there is at least one set of facts that fits the following description: (i) each of the facts in the set obtained on the evening in question; but (ii) if all the facts in the set had instead

been absent that evening then I_1 would have been absent, regardless of the stimulant level; but (iii) all of the other counterfactual dependencies depicted by the (direct) arrows in fig. 3 would have remained the same. Call any set of facts that meets this description an "actual enabler" of the role played by the stimulant on the evening in question. So my point is that there is likely at least one set of facts that actually enabled the stimulant's role on the evening in question. This it to say, information *b* is likely.

For this reason, when a reliable inquirer is presented with Mint Threat plus information *a*, she will already have a high credence that *b* is the case (at least 80 percent let's say). So, applying the same Bayesian logic as I did to information *a*, it follows: adding information *b* to Mint Threat plus information *a* will not increase a reliable inquirer's credence in P_{MINT} by very much at all. This is premise three of my overall argument.

Note, again, that my argument for premise three assumes only the Bayesian Assumption and the Enabling Assumption. But the Enabling Assumptions is about counterfactual dependence, not about causation. And so my argument for premise three holds for any view of causation at all—even views in which absence causation is impossible, or views in which facts about causation are not determined by facts about counterfactual dependence.

Premise 4: Adding Information *c*

What happens to a reliable inquirer's causal judgments when one adds information *c*? Information *c* is a little more specific than information *b*. Information *c* says that one of these actual enablers (of the stimulant's role that evening) was Joanna's having recently eaten a black coloured pill. (To keep things simple in my discussion below, all of my examples of enablers will be enablers that consist of a single fact, rather than enablers that consist of multiple facts acting in conjunction with each other.)

To consider the significance of this information *c*, compare proposition *c* with proposition *c**: one of these actual enablers was Joanna's eating a white coloured pill. And let's consider how a reliable inquirer would respond if they were to instead learn that *c** is true. One might say that we are thereby considering Joanna's eating a white pill as a putative enabler, so to speak, of the stimulant's role. Now, a reliable inquirer who knows Mint Threat (plus information *a*–*b*) would respond to learning *c** in exactly the same way that a reliable inquirer would respond to learning *c*, I assume. It is simply not relevant for judging what caused what on that particular evening to know the colour of the pill that actually enabled the stimulant's role. In this respect, Joanna eating a black pill and Joanna eating a white pill are putative enablers that a reliable inquirer will treat equivalently (for the purposes of causal inference in this setting). As a shorthand, I will say that the white pill is causally equivalent to the black pill (as a putative enabler of the stimulant's role that evening).

I assume that there are lots of other putative enablers that are causally equivalent

to the black pill. Take for example the proposition *c***: one of these actual enablers (of the stimulant's role that evening) was Joanna's receiving an injection. A reliable inquirer would also respond to learning *c*** in exactly the same way that a reliable inquirer would respond to learning *c* or *c**, I assume. In this respect, an injection is causally equivalent to the black pill (as a putative enabler of the stimulant's role that evening). And the same is true, I assume, of Joanna contracting a virus as a putative enabler, or Joanna suffering a psychological trauma as a putative enabler. One can extend this line of reasoning—from the black pill, to pills of all colours, to injections, to viruses, to psychological trauma—to find more and more putative enablers that are causally equivalent to the black pill. Therefore the information in *c* that is relevant for deciding what caused what that evening is the following information (c'): one of the actual enablers (of the stimulant's role that evening) was Joanna's eating a black pill or something causally equivalent to Joanna's eating a black pill. This raises the following crucial question:

> **Crucial Question**. Given that there was at least one actual enabler of the stimulant's role that evening, as *b* says there was, how likely is it that *c'* is true? That is to say, how likely is it that one of these actual enablers is something causally equivalent to Joanna's eating a black pill?

The answer to this question will be determined by two things. Firstly, how many actual enablers (of the stimulant's role that evening) were there likely to have been that evening? Just one, or a few, or many? The more actual enablers there were, the more likely that at least one of these enablers was something causally equivalent to Joanna's eating a black pill. I'm inclined to think that in biochemistry there are typically lots of enablers, for the reasons I've already given. Secondly, how many putative enablers are causally equivalent to Joanna's eating a black pill? The more there are, the more likely one of the actual enablers will be on this list of putative enablers that are causally equivalent to Joanna's eating a black pill. As I've already indicated, I'm inclined to think that this list is rather long. For these two reasons, I contend:

> **Many Causally Equivalent Enablers**. There are many putative enablers that are causally equivalent to Joanna's eating the black pill. That is to say, there are enough that information *c'* is likely, given information *b*.

It follows immediately from Many Causally Equivalent Enablers that, when presented with Mint Threat (plus information *a*–*b*), a reliable inquirer will already have a high credence (at least 80 percent let's say) that *c'* is the case. So one can apply the same Bayesian logic, as I did to information *a* and *b*, to show that adding information *c'* to Mint Threat plus information *a b* will not increase a reliable inquirer's credence in P_{MINT} by very much at all. Since *c'*, by definition, contains only the information in *c* that is relevant for making causal judgments, it follows that adding information *c* to Mint Threat (plus information *a*–*b*) will not increase a reliable inquirer's credence in P_{MINT} by very much at all either. That's premise four of my overall argument.

Note that Many Causally Equivalent Enablers (MCEE) does indeed make a claim about causation. It claims that there is a large list of putative enablers that should be treated equivalently for the purposes of causal inference. And it is certainly reasonable to question my contention that MCEE is true. Indeed, Section 6 will explore one way in which one might try to reject MCEE. For the moment, I just want to note that MCEE does not embody some of the more controversial assumptions that one might make about causation. For example, MCEE could hold in virtue of *c'* containing a long list of presences, rather than absences. And so MCEE is entirely compatible with the idea that absence causation is impossible. Indeed, MCEE is entirely compatible with the idea that facts about causation are not determined by facts about counterfactual dependence alone.

Premise 5: Adding Information *d–f*

What about the addition of information *d–f*? Information *d* just provides a name (hypnotase *B*) for intermediate variable I_1. And according to Mint Threat plus information *a–c*, this variable is a variable upon which sleep is positively causally dependent. So, in providing the memorable name hypnotase *B*, information *d* doesn't provide any information of additional relevance over and above the information already contained in information *a–c* and Mint Threat. And so adding information *d* to Mint Threat (plus information *a–c*) will not change a reliable inquirer's credence in P_{MINT} at all.

Moving on, information *e* just provides an alternative name for the antidote (hypnotase *A*). And, according to Mint Threat, this variable is a variable upon which sleep is positively counterfactually dependent. So in providing the memorable name hypnotase *A*, information *e* doesn't contain any information of additional relevance. And so adding information *e* to Mint Threat (plus information *a–d*) will not change a reliable inquirer's credence in P_{MINT} at all.

Moving on, information *f* doesn't contain any additional information that is relevant to evaluating causes and effects. It just provides some alternative names for Jo (Anna, Joanna) and for the Zeus mint (amber pill). And so adding information *f* to Mint Threat (plus information *a–e*) will not change a reliable inquirer's credence in P_{MINT} at all.

Premises 9 and 10: Adding Information *g–k*

Consider a reliable inquirer who knows the information that defines Redundant Pill. How will her causal judgments change when she learns information *g*? Redundant Pill already tells her that: (5) Anna will fall asleep at midnight if and only if either hypnotase *A* or hypnotase *B* are present. But I've already argued that counterfactual dependencies in biomedicine are likely mediated by long chains of counterfactual dependencies between

one hundred or more chemicals. So it is likely that there is a chemical C such that: Joanna will fall asleep at midnight if and only if C is present; and C is present if and only if either hypnotase A or of hypnotase B is present. Indeed, let C denote the the "earliest" variable in the chain of variables that mediate this counterfactual dependence between Anna's falling asleep on the one hand, and hypnotase A and B on the other. (Thus C is the first point in the chain at which the information about hypnotase A's presence is "merged", as it were, with the information about hypnotase B's presence.) But, given that chains of counterfactual dependence in biomedicine are very long, it is likely that the counterfactual dependence of sleep upon the presence of chemical C is itself mediated by an absence—for reasons I gave earlier in the section. This too is part of what I call the Mediation Assumption. In sum, it is likely that there is a variable I_2 such that: (g') Joanna will fall asleep at midnight if and only if I_2 is absent; I_2 is absent if and only if C is present; but C is present if and only if either hypnotase A or of hypnotase B is present. But *g'* entails *g*. So *g* is likely too. Therefore adding information *g* to Redundant Pill will not decrease a reliable inquirer's credence in P_{AMBER}, at least not by very much, one can show, again using a Bayesian logic. [16] This is the first part of premise nine from my overall argument.

For similar reasons, the Mediation Assumption also entails that adding information *h* to Redundant Pill (plus information *g*) will not decrease a reliable inquirer's credence in P_{AMBER}, at least not by very much. This is the second part of premise nine from my overall argument. Note that my argument for premise nine relies on the Bayesian Assumption and the Mediation Assumption alone. It is neutral with respect to the nature of causation.

What about premise ten? Adding information *i–k* to Redundant Pill (plus information *g–h*) will not alter her credence in P_{AMBER} at all. After all, information *i* just provides a name (disruptase) for a intermediate variable. And our inquirer already knows, given information *g*, that this intermediate variable is a variable upon which Joanna's falling asleep is negatively counterfactually dependent. Similarly, information *j* just provides a name (the stimulant) for a second intermediate variable. And our inquirer already knows, given information *h*, that this variable is a variable upon which Joanna's falling asleep is negatively counterfactually dependent.

6 The Challenge of Enabling Variables

I am persuaded by the foregoing argument, whose conclusion is that the typical person's causal intuitions are thoroughly unreliable. But I acknowledge that it is reasonable to question whether or not Many Causally Equivalent Enablers is true, and so it is reasonable to resist the foregoing argument. Even so, the foregoing argument presents a challenge for philosophers who want to defend the typical person's intuitions as some-

what reliable. The challenge is as follows:

1. Offer a principled answer to the question: what information about enabling (and intermediate) variables is relevant to reliable causal judgment? and conversely what information about enabling (and intermediate) variables is negligible?

2. Show that this principled answer entails that there are relatively few putative enablers that one should treat equivalently to the black pill. (Thus show that MCEE is false, and that my argument in Section 5 is unsound.)

3. Thereby show that the information *c* (that it was a black pill that enabled the stimulant's role that evening) is highly relevant information—information that warrants a "reversal" of one's initial causal judgment in Mint Threat.

That's the challenge for anyone who wants to neutralize my argument in Section 5 that the typical person's causal intuitions are thoroughly unreliable. In this section, I will suggest that it will be more difficult to meet this challenge than it might appear. In particular, I will explore the suggestion that (a) causal knowledge is norm-involving, and that (b) this normativity of causal knowledge entails that MCEE is false.

In what sense might causal knowledge be norm-involving? To know what causes what in any given case, an inquirer must first make a normative judgment, some philosophers claim. For illustration, suppose that Katie fails to water her plant, and her plant then dies. Intuitively, Katie's failure to water her plant was a cause of its death. But Queen Elizabeth's failure to water Katie's plant was, intuitively, not a cause of its death. And one knows this—to cut a long story short—because one knows that it is normal for Katie to water her plant, but it is not normal for the Queen to water Katie's plant. Normal here means "apt" or "fitting" rather than "occurs frequently". In the jargon, the default value of the Katie Watering variable is on, and the deviant value is off; whereas the default value of the Queen Watering variable is off, and the deviant value is on. In general, to know what caused what, one has to first make a normative judgment of the above sort. Many philosophers contend that our knowledge of (single-case event) causation is, in this respect, norm-involving.[17]

But if our knowledge of causation is norm-involving, then it's plausible that a reliable inquirer will respond to the following two pieces of information differently: (c) the stimulant's role was actually enabled by Joanna's eating a black pill; (c***) the stimulant's role was actually enabled by Joanna's eating a large meal. For example, it's abnormal for Joanna to eat a pill, but it is normal for Joanna to eat a large meal, one might think. In light of this, consider the class *c'* of putative enablers that are causally equivalent to the black pill. If causal knowledge is norm-involving, then perhaps class *c'* only contains putative enablers that are abnormal. If so, class *c'* may well be smaller than if causation were not norm-involving.

Nevertheless, I still contend that Many Causally Equivalent Enablers is true: there are "enough" putative enablers that a reliable inquirer will treat equivalently to the black

pill. A reliable inquirer will still treat the following abnormal putative enablers equivalently: Joanna eats a black pill, Joanna eats a white pill, Joanna receives an injection, some of Joanna's genes are knocked out using CRISPR technology, Joanna undergoes a course of psychotherapy, and so on. So, even when class *c'* is restricted to putative enablers that are abnormal, class *c'* remains a sufficiently broad class, I insist. Therefore, even if causal knowledge is norm-involving, Many Causally Equivalent Enablers (MCEE) still stands, I contend.

Of course, to say this is not to provide decisive reason to think that MCEE is true. I have no decisive reason to offer to establish MCEE. What I can do, however, is rebut the following argument whose conclusion is that MCEE is false: [18]

1. Some putative enablers of the stimulant's role are normal.
2. Joanna's eating black pill, in contrast, is a putative enabler that is abnormal.
3. When engaging in causal inference, reliable inquirers should treat each abnormal enabler the same as any other abnormal enabler, but differently from any normal enabler.
4. So each putative enabler in class c' (of putative enablers a reliable inquirer will treat equivalently to the black pill) is abnormal.
5. Abnormal enablers are much less likely to occur than normal enablers.
6. So, given that the stimulant's role was actually enabled on the evening in question, it is unlikely that one of the actual enablers belongs to class c'. That is to say, Many Causally Equivalent Enablers is false.

The problem with the above argument, I suggest, is that it trades on an ambiguity in the concept of an event being normal. If normal means "occurs with high probability or frequency", then the warrant for believing premise two is unclear. Is it unusual for Joanna to eat a black pill? Redundant Pill certainly doesn't say anything about how usual it is for Joanna to take the black and amber pills. What's more, if normal means "occurs with high probability or frequency" then premise three is unattractive. After all, most philosophers who think that causal knowledge is norm-involving are clear that the normal vs abnormal distinction is to be interpreted in terms of what is apt or appropriate, not in terms of what is probable or frequent, as I've already noted. In contrast, if normal means "is apt or appropriate", then the warrant for premise four is unclear. Why think that inapt enablers are much less likely to occur that apt enablers? So, absent further elaboration, this argument doesn't provide a strong reason to think that MCEE is false. Of course, this doesn't establish decisively that MCEE is true. But it does illustrate the work that remains to be done by philosophers who wish to defend the typical person's intuitions as reliable.

7 Conclusion

Section 4 provided a deductively sound argument for Disunity: no coherent metaphysical theory of causation can both (a) issue a decisive causal verdict about threat-and-savour cases such as Mint Threat in line with the typical person's intuitions, and (b) issue a decisive causal verdict about redundancy cases such as Redundant Pill in line with the typical person's intuitions. Philosophers who have attempted to provide a unified theory of causation that respects both these intuitions are attempting the impossible. My argument for Disunity did not rely on any controversial philosophical assumptions. It can only be resisted by performing a more rigorous empirical survey than the informal survey whose results I reported in section 2.

Turning from metaphysics to epistemology, Section 5 also provided a deductively sound argument for Defeasible: the typical person's intuitions in at least one of *a* and *b* are (at best) defeasible evidence. Again, this argument did not rely on any controversial philosophical assumptions.

Somewhat more controversially, Section 5 also provided an argument that the typical person's intuitions are not only defeasible but also thoroughly unreliable. The key assumption of this argument (MCEE) does not assume that facts about causation are fully determined by facts about counterfactual dependence. Nor does it assume that absence causation is possible. Nevertheless, the MCEE assumption is open to question. This raises what I call the challenge of enabling variables:

1. Offer a principled answer to the question: what information about enabling (and intermediate) variables is relevant to reliable causal judgment? and conversely what information about enabling (and intermediate) variables is negligible?

2. Show that this principled answer entails that there are relatively few putative enablers that one should treat equivalently. (Thus show that MCEE is false, and that my argument in Section 5 is unsound.)

3. Thereby show that information about enablers is highly relevant information—information that often warrants a "reversal" of one's initial causal judgment.

Acknowledgements and Funding

Thank you to audiences at EIPE Rotterdam and TILPS Tilburg for discussions of these ideas. And thank you Martin van Hees, Conrad Heilmann, and Fred Muller for your comments on an ancestral version of this manuscript. This work has received funding from the European Research Council under the European Unions Horizon 2020 Research and Innovation Programme, under grant agreement no 715530.

Notes

[1]Lewis (2004), who examined this theory in Lewis (1973), agrees. See also McDermott (1995), Collins, Hall, and Paul (2004), Hitchcock (2007), and Weslake (n.d.) and the papers cited therein.

[2]Hall (2000), Hitchcock (2003), Maudlin (2004, 422), Hall (2004, 230), Hall (2006), and Hitchcock (2007, 498), and Northcott (2021) for example.

[3]For this case see Hall (2000, 276) and Paul and Hall (2013, chap. 3) as well as the citations in the footnote below.

[4]I define threat-and-saviour cases as cases of the form: (i) an event C caused the presence of *Threat*; and (ii) if *Saviour* had been absent, then *Threat* would have caused E to be absent; but (iii) C also caused *Saviour*; and so (iv) *Saviour* caused E, despite the presence of *Threat*. See also the case of Careful Poisoning as discussed by Weslake (n.d., sec. 3), and see also Yablo (2004) on Stockholm Syndrome cases, as well as Kvart (1991), Hall (2000), Pearl (2000, S10.3.4), Hitchcock (2003), Halpern and Pearl (2005), Hiddleston (2005), Bjornsson (2007), Hitchcock (2007), Hall (2007), and Paul and Hall (2013, chap. 3).

[5]Firstly, I define redundancy cases as as cases of the form: (i) if event C had been absent, then some other event B would have caused event E to occur; and (ii) if B had been absent, then C would have caused E to occur. Secondly, contrast the definition of redundancy in Lewis (2004, 80); and see Won (2014) for various problems for Lewis' definition of redundancy. Thirdly, in some redundancy cases, many people have the intuition that event C "pre-empts" event B from causing event E. In other redundancy cases, in contrast, people have the intuition that B is also a cause of E.

[6]Other than the variable on the left-hand side of statement 1 itself, namely the Awake variable.

[7]In an informal survey of $n = 13$ philosophers, the mean score was 2.3, on a Likert scale of 1 'strongly disagree' and 7 'strongly agree'. The modal score was 1. In a second informal survey of $n = 37$ non-philosophers and $n = 11$ philosophers, the mean score was 2.16 and 2.55 respectively. The modal score was 1 and 1 respectively.

[8]In the first survey, the mean score was 6.0, on a Likert scale of 1 'strongly disagree' and 7 'strongly agree'. The modal score was 7. In the second survey, the mean score was 5.84 and 6.18 respectively. The modal score was 7 and 7 respectively.

[9]In the first survey, the mean score was 2.3 (again), on a Likert scale of 1 'strongly disagree' and 7 'strongly agree'. The modal score was 1. In this second study, the mean score was 1.57 and 1.73 respectively. The modal score was 1 and 1 respectively.

[10]Compare the five events *A B C D E* that occur in Boulder Threat-and-Saviour with the five events *A' B' C' D' E'* that occur in Redundant Alarm. Note that (E) the hiker walks home if and only if (B) the hiker ducks or it's not the case that (D) a boulder falls towards hiker. Analogously it's true that (E') Hanna wakes if and only if the pain receptors on her face fire or her auditory system is startled. That is to say, if and only if (B') the pain receptors on her face fire or it's not the case that (D') her auditory system fails to be startled. But it's true that (B) the hiker ducks if and only if (C) the boulder is dislodged. Analogously it's true that (B') the pain receptors on Hanna's face fire if and only if (C') the cat leaps on her face. But note that (D) a boulder falls towards hiker if and only either the boulder in question is dislodged, or a second

boulder is dislodged. That is to say, if and only either (C) the boulder in question is dislodged, or it's not the case that (A) no second boulder was dislodged. Analogously, Hanna's auditory system will be startled (at seven) if and only if her alarm is set for seven, but the cat hasn't lept on her face first. That is to say: (D') Hanna's auditory system fails to be startled at seven if and only if (C') the cat lept on her face just before seven, or it's not the case that (A') Hanna's alarm is set for seven.

[11] See Hitchcock (2007), Hall (2007), Halpern (2008), Paul and Hall (2013) for reasons to endorse *iii*.

[12] Suppose, for example, that $E_1, E_2, E_3 \ldots E_n$ is a series of events in which each event in the series is spatio-temporally contiguous to the event that immediately follows it in the series, and indeed is a cause of that following event. To say that causation is transitive is, of course, to say that each event is therefore a cause of any of the following events in the series. To say that causation is an intrinsic relation is to say that this causal chain would remain intact (each event would remain a cause of each event later on in the series) even if circumstances extrinsic to these events were to be radically different. Take for example Redundant Alarm, in which the cat's leaping E_1 caused a pain in Hanna's face E_2, and the pain E_2 in turn caused Hanna's waking E_3. Note that, even if one were to add the extrinsic event that Hanna set her alarm clock the night before to go off at 7am, E_1 would still cause E_2, and E_2 would still cause E_3, and by transitivity E_1 would still cause E_3.

[13] That's undeniable: counterfactual dependence is itself an extrinsic relation. Indeed, Hall also notes that the counterfactual conception of causation is very difficult to reconcile with the concept of causation as transitive, unless one is willing to accept some very bizarre claims about causation (Hall 2000, 2004).

[14] Although I'm discussing counterfactual dependence here, not causation, see Thomson (2003) and Schaffer (2004) and Sartorio (2009) for discussion about how causation may or may not be mediated by absences.

[15] $Cr(PA) \leq Cr(P)$. So $Cr(PA)/Cr(A) \leq Cr(P)/Cr(A)$. So $Cr(P|A) \leq Cr(P)/Cr(A)$. And so $Cr(P|A) - Cr(P) \leq [Cr(P)/Cr(A)] - Cr(P) = Cr(P)([1/Cr(A)] - 1) = Cr(P)[1 - Cr(A)]/Cr(A)$.

[16] One can prove, using the Bayesian Assumption, that if a reliable inquirer's credence in P_{AMBER} is less when presented with Redundant Pill plus information g (than it is when presented with Redundant Pill alone), then this decrease in credence will be, at most $[1 - Cr(P_{AMBER})]\dfrac{1 - Cr(G)}{Cr(G)}$. Here's the proof. $Cr(G\neg P) \leq Cr(\neg P)$. So $Cr(G\neg P)/Cr(G) \leq Cr(\neg P)/Cr(G)$. So $Cr(\neg P|G) \leq Cr(\neg P)/Cr(G)$. And so $Cr(\neg P|G) - Cr(\neg P) \leq [Cr(\neg P)/Cr(G)] - Cr(\neg P) = Cr(\neg P)([1/Cr(G)] - 1) = Cr(\neg P)[1 - Cr(G)]/Cr(G)$. And so $[1 - Cr(P|G)] - [1 - Cr(P)] \leq Cr(\neg P)[1 - Cr(G)]/Cr(G)$. And so $Cr(P) - Cr(P|G) \leq Cr(\neg P)[1 - Cr(G)]/Cr(G)$.

[17] See Menzies (2004), McGrath (2015), Halpern (2008), Hall (2007), Hitchcock (2007), Hitchcock and Knobe (2009), and Gallow (2021) for this norm-involving approach to causation.

[18] Many thanks to an anonymous referee for drawing my attention to something like this argument.

References

Alexander, Joshua, and Jonathan M. Weinberg. 2008. "Analytic Epistemology and Experimental Philosophy." *Philosophy Compass* 2: 56–80. https://doi.org/10.1111/j.1747-9991.2006.00048.x.

Bjornsson, Gunnar. 2007. "How Effects Depend on Their Causes, Why Causal Transitivity Fails, and Why We Care about Causation." *Philosophical Studies* 133: 349–90. https://doi.org/10.1007/s11098-005-4539-8.

Collins, John, Ned Hall, and L. A. Paul. 2004. "Counterfactuals and Causation: History, Problems, and Prospects." In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul, 1–58. Cambridge MA: MIT Press.

Gallow, J. Dmitri. 2021. "A Model–Invariant Theory of Causation." *Philosophical Review* 130: 45–96. https://doi.org/10.1215/00318108-8699682.

Hall, Ned. 2000. "Causation and the Price of Transitivity." *Journal of Philosophy* 97: 198–222. https://doi.org/10.2307/2678390.

———. 2004. "Two Concepts of Causation." In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul, 225–76. Cambridge MA: MIT Press.

———. 2006. "Philosophy of Causation: Blind Alleys Exposed; Promising Directionshighlighted." *Philosophy Compass* 1: 86–94. https://doi.org/10.1111/j.1747-9991.2006.00002.x.

———. 2007. "Structural Equations and Causation." *Philosophical Studies* 132: 109–36. https://doi.org/10.1007/s11098-006-9057-9.

Halpern, Joseph Y. 2008. "Defaults and Normality in Causal Structures." Https://arxiv.org/abs/0806.2140.

———. 2016. *Actual Causality*. Cambridge MA: MIT Press. https://doi.org/10.7551/mitpress/9780262035026.001.0001.

Halpern, Joseph Y., and Christopher Hitchcock. 2015. "Graded Causation and Defaults." *British Journal for the Philosophy of Science* 66: 413–57. https://doi.org/10.21236/ada582589.

Halpern, Joseph Y., and Judea Pearl. 2005. "Causes and Explanations: A Structural-Model Approach. Part i: Causes." *British Journal for the Philosophy of Science* 56: 843–87. https://doi.org/10.1093/bjps/axi147.

Hiddleston, Eric. 2005. "Causal Powers." *The British Journal for the Philosophy of Science* 56: 27–59. https://doi.org/10.1093/phisci/axi102.

Hitchcock, Christopher. 2003. "Of Humean Bondage." *The British Journal for the Philosophy of Science* 54: 1–25. https://doi.org/10.1093/bjps/54.1.1.

———. 2007. "Prevention, Preemption, and the Principle of Sufficient Reason." *The Philosophical Review* 116: 495–532. https://doi.org/10.1215/00318108-2007-012.

———. 2009. "Structural Equations and Causation: Six Counterexamples." *Philosophical Studies* 144 (3): 391–401. https://doi.org/10.1007/s11098-008-9216-2.

Hitchcock, Christopher, and Joshua Knobe. 2009. "Cause and Norm." *Journal of Philosophy* 106: 587–612. https://doi.org/10.5840/jphil20091061128.

Ingthorsson, Rognvaldur. 2007. "Is There a Problem of Action at a Temporal Distance?" *SATS Northern European Journal of Philosophy* 8: 138–54. https://doi.org/10.1515/sats.2007.138.

Kvart, Igal. 1991. "Transitivity and Preemption of Causal Impact." *Philosophical Studies* 64: 125–60.

Lewis, David K. 1973. *Counterfactuals*. Oxford: Blackwell.

———. 1983. "Postscript to Causation." In *Philosophical Papers*, 172–213. Oxford: Oxford University Press.

———. 2004. "Causation as Influence." In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul, 75–106. Cambridge MA: MIT Press.

Machery, Edouard. 2017. *Philosophy Within Its Proper Bounds*. Oxford: Oxford University Press.

Machery, Edouard, and Elizabeth O'Neill. 2014. *Current Controversies in Experimental Philosophy*. Abingdon: Routledge–Taylor.

Maudlin, Tim. 2004. "Causation, Counterfactuals and the Third-Factor." In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul, 419–44. Cambridge MA: MIT Press.

McDermott, Michael. 1995. "Redundant Causation." *British Journal for the Philosophy of Science* 46: 523–44.

McGrath, Sarah. 2015. "Causation by Omission: A Dilemma." *Philosophical Studies* 123: 125–48. https://doi.org/10.1007/s11098-004-5216-z.

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

———. 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, 153–74. Oxford: Oxford University Press.

Northcott, Robert. 2021. "Pre-Emption Cases May Support Not Undermine the Counterfactual Theo-

ryof Causation." *Synthese* 198: 537–55.
Paul, L. A., and Ned Hall. 2013. *Causation: A User's Guide*. Oxford: Oxford University Press.
Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
Rose, David. 2017. "Folk Intuitions of Actual Causation: A Two-Pronged Debunking Explanation." *Philosophical Studies* 174: 1323–61. https://doi.org/10.1007/s11098-016-0762-8.
Sartorio, Carolina. 2009. "Omissions and Causalism." *Nous* 43: 513–30. https://doi.org/10.1111/j.1468-0068.2009.00716.x.
Schaffer, Jonathan. 2004. "Causes Need Not Be Physically Connected to Their Effects: The Case for Negative Causation." In *Contemporary Debates in Philosophy of Science*, edited by Christopher Read Hitchcock, 197–216. Oxford: Blackwell.
Thomson, Judith Jarvis. 2003. "Causation: Omissions." *Philosophy and Phenomenological Research* 66: 81–103. https://doi.org/10.1111/j.1933-1592.2003.tb00244.x.
Weslake, Brad. n.d. "A Partial Theory of Actual Causation." *British Journal for the Philosophy of Science*.
Won, Chiwook. 2014. "Overdetermination, Counterfactuals, and Mental Causation." *Philosophical Review* 123: 205–29. https://doi.org/10.1215/00318108-2400566.
Yablo, Stephen. 2004. "Advertisement for a Sketch of an Outline of a Proto-Theory of Causation." In *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul, 119–38. Cambridge MA: MIT Press. https://doi.org/10.1093/acprof:oso/9780199266487.003.0005.