Against The Causal Independence Principle for Counterfactuals

Joshua Edward Pearson, ANU
joshuaedwardpearson@gmail.com
AAP 2026
6/7/26

I. The Causal Independence Principle (CIP)

Breaking News! Colin and Frank are both in their offices. Colin flipped a fair coin and it landed heads. Meanwhile, Frank didn't scratch his nose. Assuming Colin and Frank's offices are causally isolated, the following seems true:

I adapt the case from Edgington (2003), which, though it didn't feature Colin, featured 'Frank in Australia'.

(1) Had Frank scratched his nose, Colin's coin still would have landed heads.

Counterfactuals like (1) are often called 'Morgenbesser counterfactuals': Slote (1978) credits Sydney Morgenbesser with observing the importance of them.

This and similar examples motivate:

Causal Independence Principle (CIP)
If A and C are true, and the mechanisms that settle whether A and whether C are causally independent from one another, the counterfactual $\neg A \,\Box\!\!\rightarrow C$ is true.

Pronounce 'CIP' like 'kip'.

I use '$\Box\!\!\rightarrow$' merely as shorthand for writing out counterfactual conditionals in English.

CIP is extremely popular. I'm going to argue it's false.

E.g. Edgington (2003) and Schaffer (2004) use principles like CIP to criticise Lewis (1973, 1979); and influential causal modelling approaches — e.g. Galles and Pearl (1998) — are essentially built around it.

- I'll start in **II** by giving a simple, not-so-novel and less persuasive argument against CIP from considerations about chance.
- In **III** I offer an improved argument, showing how endorsing CIP conflicts with attractive logical principles for counterfactuals.
- In **IV**, I'll argue CIP seems true because counterfactual truth depends on which facts context tells us to hold fixed. While context often requires keeping facts causally independent from the antecedent fixed, it need not.

Hájek (2026a) and Goodsell (2022) offer different arguments against CIP; however, my anti-CIP picture is quite different to theirs. Similarly to me, Briggs (2012) and Santorio (2019) outline ways in which a CIP-vindicating semantics may be logically revisionary. They do so concerning specific causal-modelling approaches to counterfactuals; in this respect, my arguments are more general.

II. Causal Independence, Chance and Mights

Suppose Alice does not flip her fair, objectively chancy coin at time t. The following 'might-counterfactual' sounds true:

Alice's coin is 'objectively chancy' in that the laws and history of the world never fully settle how her coin will land if flipped. All of the coins we consider will be objectively chancy in this way.
I assume this kind of indeterminism (about laws) for simplicity; I suspect my arguments are consistent with determinism, but I leave fleshing this out in full detail to another time.

(2) If Alice had flipped her coin, it might have landed heads.

Further, (2) seems true in an especially robust sense. To contrast, suppose I placed a bet on Alice's coin landing tails if flipped, but I forgot whether I bet on heads or tails. I can truly say:

(3) If Alice had flipped her coin and it had landed heads, I might have won my bet.

Still, I would deny (3) *if I had more information about the history of the world*. (2) remains true even supposing we know all historical facts.

I'll use '$\diamond\!\!\rightarrow$' as shorthand for a counterfactual conditional involving this special sense of 'might have'. It seems as though most cases involving objective chance can be used to derive them:

I intend to remain neutral on what this 'special sense' amounts to exactly.

High-Chance-to-Might (HCtM)
If, just before it was settled whether A, there was a substantial chance, conditional on A, that C, then $A \diamond\!\!\rightarrow C$.

Why assume a *substantial* chance, and not merely a non-zero chance? Because: though some people will want to deny that just *any* relevant chance is enough to guarantee true might-counterfactuals—in order to avoid the scepticism of Hájek (2026b)—they will likely still be attracted to this weaker principle. In fact, I think even this weaker principle can lead to sceptical problems, but that's an issue for another day.

However, HCtM causes problems for CIP. Consider:

Two Coins. Alice does not flip her fair coin at time t. Completely independently, Carol flips her fair coin at t and it lands heads.

Here, CIP gives us (4) and HCtM gives us (5):

(4) If Alice had flipped her coin (at t), Carol's coin would still have landed heads.

(5) If Alice had flipped her coin, Carol's coin might not have landed heads.

(4) and (5) feel deeply in tension. Here's one explanation as to why:

Duality
$A \,\Box\!\!\rightarrow C$ and $A \diamond\!\!\rightarrow \neg C$ are inconsistent.

If Duality is right, **Two Coins** straightforwardly shows HCtM and CIP are inconsistent. But Duality is controversial. An alternative:

Indeterminacy
$A \diamond\!\!\rightarrow \neg C$ implies that $A \,\Box\!\!\rightarrow C$ is not determinately true.

I'll remain neutral on exactly what determinate truth amounts to, assuming only that it requires truth in all 'cases', allowing different accounts of what a 'case' is. I prefer to think of truth in all cases as being probability 1, likening indeterminacy of counterfactuals with indeterminacy of the future. But there are alternatives: a 'case' is how the world world be under an 'admissible precisification' of a counterfactual, or a situation consistent with what a near-omniscient observer knows—as Stalnaker (1981) suggests.

CIP is presumably a determinate truth, as is the fact that Alice's and Carol's coins are causally independent. So (4) is not just true by CIP, but determinately so. Yet HCtM predicts (5).

CIP+HCtM is a bad combo regardless of whether we opt for Duality or Indeterminacy. Is this a good argument against CIP?

Not really. The CIP-lover can just reject HCtM and instead accept:

A move like this is suggested by Edgington (2003).

*High-Chance-to-Might** (HCtM*)
If, just before it was settled whether A, there was a substantial chance, conditional on *A and S, where S conjoins all facts causally independent of whether A*, that C, then $A \diamond\!\!\rightarrow C$.

HCtM* still predicts (2) but it does not predict (5): Carol's coin is causally isolated from Alice's. And the chance, conditional on Alice's coin being flipped *and Carol's coin landing heads*, that Carol's coin lands tails, is not substantial, but 0.

(2) If Alice had flipped her coin, it might have landed heads.

III. Antecedent Specificity and Reasoning By Cases

Here are two impeccable logical principles for counterfactuals:

MIGHT-BY-CASES
If $A\&B \diamondto C$ and $A\&\neg B \diamondto C$, then $A \diamondto C$.

WOULD-BY-CASES
If $A\&B \boxright C$ and $A\&\neg B \boxright C$, then $A \boxright C$.

Our prior credence in these principles should be high; something is seriously amiss with a theory of counterfactuals that falsifies them.

Problem: I'll argue endorsing CIP means rejecting at least one of them. In particular, endorsing DUALITY requires rejecting MIGHT-BY-CASES, and denying DUALITY requires rejecting WOULD-BY-CASES.

These 'by-cases' principles follow from more general principles. E.g. WOULD-BY-CASES follows from

DISJUNCTION
If $A \boxright C$ and $B \boxright C$ then $(A \vee B) \boxright C$.

The majority of proposed logics for counterfactuals endorse DISJUNCTION, such as every single logic considered in Bernston (2021).

DUALITY
$A \boxright C$ and $A \diamondto \neg C$ are inconsistent.

Ironically, this argument itself is an instance of reasoning by cases!

III.1 Assuming Duality

Consider a new case:

Three Coins. Alice, Billy and Carol each flip a coin at time t in conditions causally independent from one another. Alice's and Carol's coins land heads; Billy's lands tails.

Three Coins is adapted from cases by Pollock (1981) and Boylan and Schultheis (2021).

CIP predicts (6).

(6) If Alice's and Billy's coins had landed the same way (as each other), Carol's coin would still have landed heads.
$(Alice = Billy) \boxright H_C$

By DUALITY, (6) entails:

(7) It's not the case that: if Alice's and Billy's coins had landed the same way, Carol's coin might have landed tails.
$\neg[(Alice = Billy) \diamondto T_C]$

However, HCtM* predicts (8) and (9):

(8) If everyone's coins had landed the same way, Carol's coin might have landed tails.
$(Alice = Billy)\&(Billy = Carol) \diamondto T_C$

(9) If Alice's and Billy's coins had landed the same way, but Carol's differently, Carol's coin might have landed tails.
$(Alice = Billy)\&(Billy \neq Carol) \diamondto T_C$

(8), (9) and (7) are a direct counterexample to MIGHT-BY-CASES!

Further, note that given DUALITY, MIGHT-BY-CASES is equivalent to:

ANTECEDENT SPECIFICITY
If $A \boxright C$ then ($A\&B \boxright C$ or $A\&\neg B \boxright C$).

ANTECEDENT SPECIFICITY must be denied, too. (6) and DUALITY applied to (8) and (9) provide a counterexample. This more bad news, and it sets up the problem if the CIP-lover denies DUALITY.

To see the plausibility of ANTECEDENT SPECIFICITY, consider the contrapositive. If it's neither the case that C would be true were $A\&B$ true, nor that C would be true were $A\&\neg B$ true, then it feels impossible to see how it could be that C would be true were A true.

III.2 Denying Duality

The problem in this case stems from:

Determinacy Closure
If p is determinately true, and p entails q, then q is determinately true.

This is plausibly true regardless of our interpretation of determinacy.

Note (6) is, if true, not merely true but determinately so. By Determinacy Closure, this means Antecedent Specificity is only true if the following disjunction is also determinately true:

(6) If Alice's and Billy's coins had landed the same way, Carol's coin would still have landed heads. $(Alice = Billy) \boxright H_C$

After all, (6) can be derived from CIP—which is presumably determinate if true—and that the coins in **Three Coins** are causally independent, which we can stipulate is a determinate truth.

(10) $[(Alice = Billy) \& (Billy = Carol) \boxright H_C]$; or
$[(Alice = Billy) \& (Billy \neq Carol) \boxright H_C]$

But (10) is not determinately true. Under any reasonable theory of indeterminacy here—e.g. probability/admissible precisifications/idealised observers—there's a case in which Carol's coin would land tails in both the scenario in which all the coins land the same way and the scenario in which Carol's coin lands differently from Alice's and Billy's.

(10) in English: either Carol's coin would have landed heads if all of their coins had landed the same way, or it would have landed heads if Carol's coin had landed differently to both Alice's and Billy's.

So, Antecedent Specificity must still be denied. But it gets worse. Opponents to Duality—and proponents of Indeterminacy— tend to accept:

Conditional Excluded Middle (CEM)
Either $A \boxright C$ or $A \boxright \neg C$;
Equivalently, if $\neg(A \boxright C)$, then $A \boxright \neg C$.

This is for a few reasons. One is that, given Indeterminacy, apparent counterexamples to CEM, an otherwise attractive principle, can be explained away. On this view, what might seem like cases in which neither $A \boxright C$ nor $A \boxright \neg C$ are treated just cases in which neither are determinately true.

The classic example the pair: *If Bizet and Verdi had been compatriots, they would have been Italian* and *If Bizet and Verdi had been compatriots, they would have been French.*
See, especially, Stalnaker (1981).

The kicker is that, given CEM, Would-by-Cases is equivalent to Antecedent Specificity, and so must be rejected!

More concretely, the problem for Would-by-Cases is that (10) is not a determinate truth. So, in some case, c, Carol's coin both would land tails if all the coins land the same way and if Carol's coin lands differently to Alice's and Billy's. If Would-by-Cases holds, it holds in all cases. Applying it c means that, in c, Carol's coin would land tails if Alice's and Billy's coins landed the same way—contrary to (6), which was meant to be determinately true!

We can summarise my argument against CIP as follows:

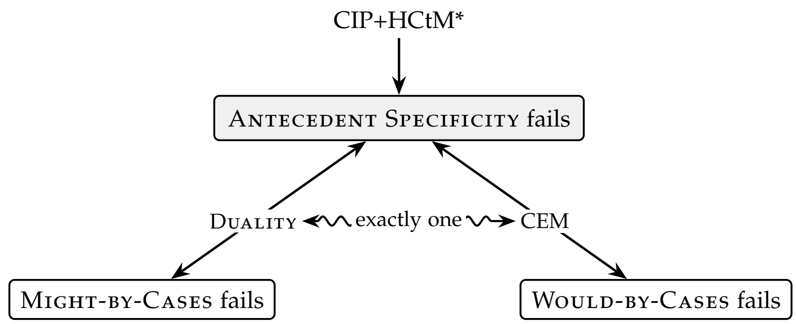

Figure 1. CIP+HCtM* forces us to give up Antecedent Specificity. But Antecedent Specificity is equivalent to Might-by-Cases given Duality, and to Would-by-Cases given CEM, and usually exactly one of Duality or CEM is endorsed (the squiggle).

Two alternative responses to my argument. (i) deny *both* Duality and CEM. This is interesting but I have little idea what it looks like. One option is to follow Edgington (1995) and insist counterfactuals have no truth values (and so no truth-preservation principles). However, I'm unsure whether this would avoid all forms of the arguments I've given here.

(ii) Bite the bullet on denying Would-by-Cases, but insist it is nonetheless a *reasonable inference* (see Schultheis (2026))—anyone who knows the premises of Would-by-Cases is in a position to know the conclusion. I am tempted by this but am not sure I prefer it to the contextualist approach below.

IV. A Contextualist Alternative

My arguments against CIP have so far been quite theoretical. Do we have any concrete counterexamples? Actually, yes. And we've essentially already seen them. Recall that, since Carol's coin is causally independent from Alice's and Billy's, CIP tells us that (6) is true.

(6) If Alice and Billy's coins had landed the same way, Carol's coin would still have landed heads. $(Alice = Billy) \mathrel{\Box\!\!\rightarrow} H_C$.

However, it's possible to situate (6) in longer conversations about **Three Coins** in which it starts to sound awkward.

David: If all their coins had landed the same way, how do you think Carol's coin would have landed?

Eliza: Well, in that case, the coins might have all landed tails, and so Carol's coin might have landed tails.

David: I agree. # Though, Carol's coin wouldn't have landed tails if Alice and Billy's coins had landed the same way.

Eliza: Really? I thought you just agreed that Carol's coin might have landed tails?

David: Not exactly. I agreed that *if* all the coins landed the same way, Carol's coin might have landed tails. # Still, Carol's coin wouldn't have landed tails if Alice's and Billy's coins had landed the same way.

The point: both of David's marked assertions are equivalent to (6). But David sounds incoherent. This is hard to explain if CIP is true.

I suspect that there's nothing much more informative to say about the relationship between causal independence and counterfactuals than the following:

- If you hear $A \mathrel{\Box\!\!\rightarrow} C$ 'in the wild', there's a default assumption that known facts causally independent of A are held fixed.
- But this default assumption is sometimes defeated when the possibility of facts causally independent of A changing is brought to salience—this is what happens in the above dialogue.

- In general, I am sympathetic to broad contextualist account on which, in any particular context c, there is a set S of facts being held fixed, such that $A \mathrel{\Box\!\!\rightarrow} C$ requires C to be true in some selected A-and-S-world. Often S contains facts causally independent of the antecedent, but it need not.

I intend these broad remarks to be compatible with various ways of fleshing out contextualism about counterfactuals, such as between e.g. Lewis (2016), who is sympathetic to DUALITY, or e.g. Schultheis (2023), who is not.

The upshot? Causal independence offers us no deep insight into counterfactuals (and vice versa) other than as a decent heuristic. Counterfactuals are often—in many contexts—used in ways that track many facts about causal independencies, but they do not track all such facts, and it's very easy to slip into contexts in which various facts about causal independencies are not tracked at all.

Contrast with Hájek (2026a), who (I believe) is committed to denying that causal independence even offers a good heuristic for counterfactual truth. Only when C had chance 1 at t conditional on A—which can happen e.g. under determinism (about laws), or when C is settled before A—is a counterfactual with causally independent A and C true. While both our views are unfriendly to CIP, mine is more friendly to CIPy intuitions.

References

Bernston, D. (2021). Near closeness and conditionals. Manuscript.

Boylan, D. and Schultheis, G. (2021). How strong is a counterfactual? *Journal of Philosophy*, 118(7):373–404.

Briggs, R. (2012). Interventionist counterfactuals. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 160(1):139–166.

Edgington, D. (1995). On conditionals. *Mind*, 104(414):235–329.

Edgington, D. (2003). Counterfactuals and the benefit of hindsight. In Dowe, P. and Noordhof, P., editors, *Cause and Chance: Causation in an Indeterministic World*. Routledge.

Galles, D. and Pearl, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1):151–182.

Goodsell, Z. (2022). Tossing Morgenbesser's coin. *Analysis*, 82(2):214–221.

Hájek, A. (2026a). "If you had caught that plane, you would be dead by now": On Morgenbesser counterfactuals. Manuscript.

Hájek, A. (2026b). *Most Counterfactuals Are False*. Manuscript.

Lewis, D. (1973). *Counterfactuals*. Blackwell.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, 13(4):455–476.

Lewis, K. S. (2016). Elusive counterfactuals. *Noûs*, 50(2):286–313.

Pollock, J. L. (1981). A refined theory of counterfactuals. *Journal of Philosophical Logic*, 10(2):239–266.

Santorio, P. (2019). Interventions in premise semantics. *Philosophers' Imprint*, 19.

Schaffer, J. (2004). Counterfactuals, causal independence and conceptual circularity. *Analysis*, 64(4):299–309.

Schultheis, G. (2023). Counterfactual probability. *Journal of Philosophy*, 120(11):581–614.

Schultheis, G. (2026). Reasonable inferences for counterfactuals. Manuscript.

Slote, M. A. (1978). Time in counterfactuals. *The Philosophical Review*, 87(1):3–27.

Stalnaker, R. (1981). A defense of conditional excluded middle. In Harper, W. L., Stalnaker, R., and Pearce, G., editors, *Ifs*, pages 87–104. D. Reidel.