Copy; Paste; Counterfactual

On David Lewis' account, counterfactual claims are analyzed by examining the nearest possible world in which an indeterministic miracle creates the nonoccurrence of the event. In that possible world, the event is altered by an indeterministic miracle, creating the nonoccurrence of that event. I aim to improve Lewis' account by way of a minor modification; Lewis appears to require a transitionary period between this indeterministic miracle and the alteration of the event in question. However, it seems that truly maximizing similarity, at least in a significant subset of cases, requires the removal of Lewis' transitionary period. In other words, the indeterministic miracle need not occur before the change in event—it can simply *be* the change in event. Creating this change maximizes similarity between the actual and possible worlds while, in many cases, avoiding Lewis' worries about downstream dissimilarities or the necessity of additional 'clean-up miracles.'

We often wonder whether, had a certain event been different, a later, seemingly connected outcome would have been different as well. Counterfactual questions appear, not only throughout philosophical debates, but in commonplace thought and other areas of academic study. If Martin Luther King had died after being stabbed in 1958, how would the racial equality movement have unfolded?[1] If the British Empire had provided the U.S. colonies greater governmental representation, would the colonies have rebelled? If the universe's gravitational constant were 0.01% greater, how would the structures of galaxies be affected?

The philosophical literature on causation and counterfactuals has been shaped by David Lewis' 1979 article, "Counterfactual Dependence and Time's Arrow." There, Lewis provides an account of counterfactual claims and how to analyze their truth value. His model has been exported to and largely adopted by various sub-fields of philosophy.

Lewis' method of counterfactual analysis contains a few basic components. First, we assume a fixed past leading up to the event in question—the event we are thinking may have gone differently. The fixed timeline continues after the event to the outcome we are interested in exploring—the outcome that might have gone differently had the event gone differently. To analyze this question, on Lewis' account, we posit a similar possible world in which a small, indeterministic miracle—"a violation of the laws of nature"—prior to the event in question creates the nonoccurrence of that event in the nearby possible world (Lewis 1979, 468). The connection between the (non)event and the outcome in question in the nearby possible world informs us about the causal connection between the two in our actual world. But it is critical that we consider the most *similar* nearby possible world to ensure a closeness of fit in our counterfactual analysis. (This should make intuitive sense; if we examine a change in events in a possible world that is radically different than our own, this will not illuminate the causal connections of our actual world.

[1] Example from Byrne (2005, 1).

Similarity secures the viable transfer of statements between worlds.) On Lewis' account, a counterfactual claim, '*If E then O*,' is true iff the closest possible world(s) hold that if E is true, then O is true, and there is not an equally similar world where E is true but O is false (1979, 465).

I aim to improve Lewis' account by way of a minor modification. Lewis appears to *require* a transitionary period in the nearby possible world between the indeterministic miracle and the alteration of the event in question.[2] In his analysis, the indeterministic miracle *leads up to* the relevant change in event. Consider the following from Lewis: "The deterministic laws of w_0 are violated at w_1 in some simple, localized, inconspicuous way. A tiny miracle takes place. Perhaps a few extra neurons fire in some corner of Nixon's brain. As a result of this, Nixon presses the button" (1979, 468). Here we can see a transitionary period between the miracle (neurons firing) and the change in event (moments later pressing a button).

Notice Lewis' emphasis on the miracle occurring "in some *simple*, *localized*, *inconspicuous* way." He emphasizes this slightness of change because we must maximize the similarity between the actual and possible world. However, it seems that truly maximizing similarity, at least in a significant subset of cases, requires the removal of Lewis' transitionary period. In other words, the indeterministic miracle need not occur before the change in event—it can simply *be* the change in event. Creating this change maximizes similarity between the actual and possible world while, in many cases, avoiding Lewis' worries about downstream dissimilarities or the necessity of additional 'clean-up miracles.' If this is correct, our account of counterfactual analysis need not include a transitionary period between miracle and event. We can then analyze many counterfactual questions by a 'copy; paste' method: creating a copy of the actual world, miraculously changing the event in question, and then examining how doing so alters the later outcome in question by pressing 'play' on both worlds.

Section 1: Removing the Transitionary Period

Say that in the real world, w_0, an event E occurs and later, a supposed outcome of that event O occurs. We want to know about the connection between E and O in our world, w_0. Specifically, we want to know if E had not occurred, whether O would have occurred. To answer this question, according to Lewis, we proceed with the following four basic components:

1. A fixed, deterministic past leading up to event E which was followed by outcome O in the actual world w_0.
2. We then posit a possible world, w_*, in which an indeterministic miracle prior to E creates the nonoccurrence of that event, ~E, in w_*.
3. The connection between E and O in the actual world will be understood by an examination of what follows concerning E vs ~E in the closest or most similar world(s).

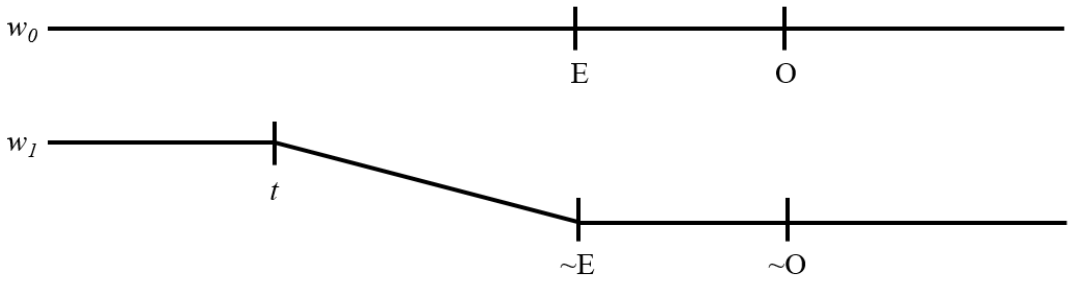
[2] His article does not speak of the possibility of a counterfactual analysis done without this transition period.

4. The counterfactual claim, "If it were true that ~E, then it would be that ~O," ~E→~O, is true for w_0 iff the closest possible worlds hold that if ~E is true, then ~O is true and there is not an equally close possible world where ~E is true but ~O is false.[3]

My main concern surrounds component two: the transitionary period between the deterministic miracle and the event of interest, E. Lewis lays out this scheme as follows:

> Until shortly before *t*, w_1 is exactly like w_0. The two match perfectly in every detail of particular fact, however minute. Shortly before *t*, however, the spatio-temporal region of perfect match comes to an end as w_1 and w_0 begin to diverge. The deterministic laws of w_0 are violated at w_1 in some simple, localized, inconspicuous way. A tiny miracle takes place (1979, 468).

We can illustrate Lewis' formulation as follows:



The actual world, w_0, shows events as they occurred: event E being followed by outcome O. We want to know their counterfactual relationship, and so in w_1, a miracle occurs at time *t* prior to E such that it does not occur, deviating from the original timeline. If O does not occur, as is shown in w_1, then the counterfactual claim "If E had not occurred, O would not have occurred" is true. (Depending on the details of what ~O looks like, it may also be accurate to frame the statement in more positive terms: "If E had not occurred, X would have occurred," if X is just to provide a positive account of ~O.)

Lewis argues that the transitionary period between the miracle and ~E in w_*, while necessary to create the divergence between worlds, should be as short as possible to (i) avoid large or many violations of natural laws between worlds, (ii) maximize similarity between worlds, and (iii) secure similarity of particular facts—the facts relevant to the counterfactual question at hand (1979, 468-472).

But there is something odd about this rationale—more specifically, about the way it is applied to Lewis' formulation of counterfactual analysis. If the transitionary period should be set up to maximize similarity across worlds, then it should not only be as short as possible; it should *disappear*, if possible. A possible world in which there is no transitory period between miracle and

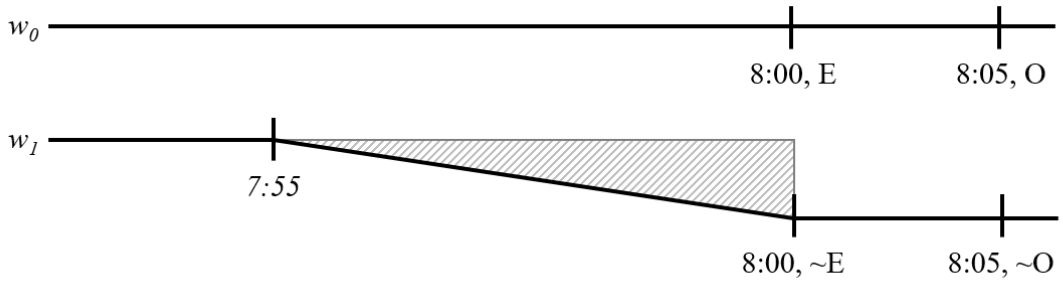
[3] These claims are modified from Lewis (1979, 465). My formulation of 4, while faithful to Lewis (1979), takes stylistic inspiration from Stalnaker (1968).

event is maximally similar world to our own. W_**'s* past would match w_0 perfectly up until the change in event, securing maximal similarity between worlds and creating the most accurate counterfactual analysis. My aim is to defend this alteration; arguing that a transitionary period is not required in counterfactual analysis.

To illustrate this change, consider a counterfactual question about a well-balanced breakfast, raised by Adam Elga: "At 8:00, Gretta cracked open an egg onto a hot frying pan. [Is] the following counterfactual true?...If Gretta hadn't cracked the egg, then at 8:05 there wouldn't have been a cooked egg on the pan" (200, 314).[4] Five minutes seems a bit long for an over-easy egg, but Elga's evaluation of this question follows Lewis' account to perfection:

> To answer, we must first ask, of the *no-crack worlds* (worlds in which Gretta doesn't crack the egg), which one is *closest* (i.e., which one is most similar to the actual world)? In order for [the above counterfactual statement to be true], it has to turn out that the closest no-crack world is one in which (A) history before 8:00 is almost exactly like actual history before 8:00, and (B) history after 8:00 differs significantly from actual history after 8:00 (2000, 314).

Here are the facts: at 8:00 in the actual world, Gretta cracks an egg into a pan (event E). At 8:05, there is a cooked egg in the pan (outcome O). We must construct a possible world in which E does not occur to examine if O would then fail to occur. To do so, as Elga states, we change the history of the actual world just before 8:00 to create ~E: the non-cracking of the egg. Then we examine how the worlds differ from that point on. Elga provides a possible miracle, occurring five minutes before the egg cracking: "If Gretta hadn't cracked the egg, then at 7:55 she wouldn't have taken an egg out of her refrigerator" (2000, 314). This minor miracle—the prevention of Gretta's removing an egg from her fridge—would prevent her from cracking it into the pan (~E) and, presumably, preventing a cooked egg from being in the pan at 8:05 (~O). From this description, we can construct a timeline for this counterfactual analysis, following Lewis' instructions:[5]



[4] Elga's point is different than mine, but his example provides a helpful way to examine Lewis' analysis of counterfactual claims.

[5] I did not make the timeline strictly to scale to allow for a closer look at the time between the miracle and the event in question (7:55 to 8:00).

Here we can see that the miracle at 7:55 creates a change in the event in question—in this case, the cracking of an egg. Since it seems that at 8:05, there is no cooked egg in the pan, we can conclude, on the assumption that w_1 is indeed the closest possible world to w_0, that Elga's counterfactual claim was true. But this assumption seems misguided. The triangular shaded area represents the difference between each timeline created by the minor indeterministic miracle (the alteration of Gretta's taking an egg out of the fridge). And it seems clear that the shaded area could be made smaller, thus increasing the similarity between possible and actual worlds, *and* increasing the accuracy of the counterfactual analysis.

If we could move the miracle forward, say to 7:56, that would shrink the shaded area of difference, increasing the similarity between worlds. Moving the miracle forward to 7:57 would further increase similarity. We can run this line of reasoning down to a moment before the event—down to the smallest fraction of time possible before the event. Shrinking the transitionary period makes the nearby possible world more and more similar, rendering the counterfactual analysis more and more accurate. So why not remove the transitionary period altogether?

I suggest we should. I contend that the *most* similar world, and therefore the world which will yield the most accurate counterfactual analysis, should be examined without a miracle prior to the event; instead, the miracle should simply *be* the change in event itself. Our analysis would then function by 'pausing' the actual world at event E, making a copy of it, swapping the detail relevant to the question at hand—in this case ~E—and then 'pressing play' on both worlds, examining differences with respect to O.

Section 2: Case Studies

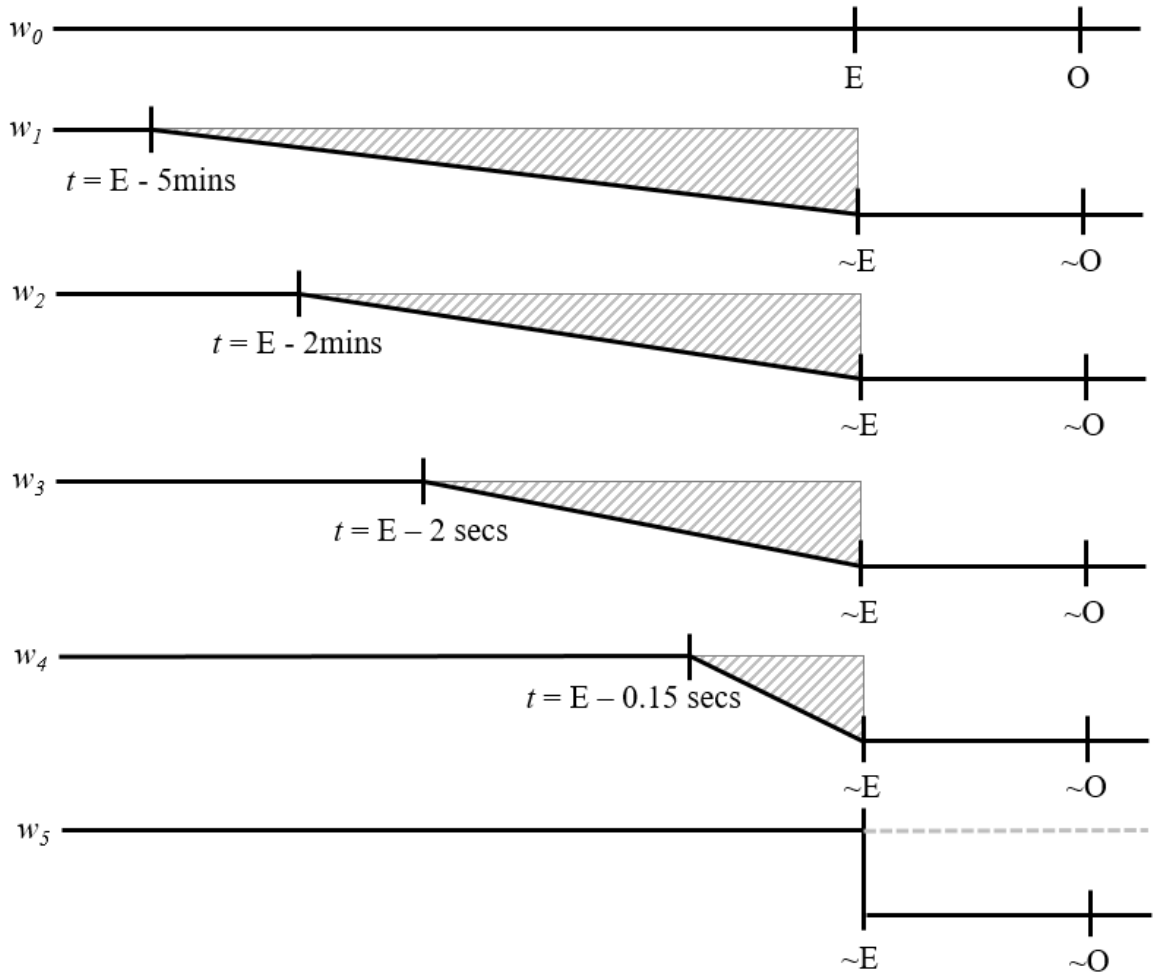
One may worry that my suggestion violates other desiderata—specifically, the requirement to minimize many violations of laws between the actual and possible worlds. Creating a change at the moment of the event may create the need for many subsequent 'clean-up' miracles to secure similarity of particular facts between the actual and possible world. For instance, if the timelines between w_0 and w_1 were identical up until the moment Gretta cracked an egg into the pan, and then in w_1, the egg was suddenly removed, this would create many issues. Gretta would be left incredulous as she cracked an egg over her hot pan, only for the egg to disappear as it hit its surface; she would have the memory of retrieving the egg from the fridge and cracking it over the bowl without the egg being in the pan as it should be. These immediate oddities would presumably propagate forward, creating downstream differences in w_1 that had nothing to do with the outcome in question (the presence of a cooked egg in the pan at 8:05). To prevent these pollutions would require many 'clean-up' miracles: her memory of the past five minutes would have to be wiped or altered, the pan made cool, the stove turned off, etc. But this worry is of particular concern, as "it is of the *first* importance to avoid big, widespread, diverse violations of law [and only of] *second* importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails" (Lewis 1979, 472).

However, I am not sure that objections of this kind hold for many cases. And because I am aiming to remove the *necessity* of a transitionary period in Lewis' account of counterfactual analysis, if I can show that cases exist which do not succumb to these worries, then I have succeeded in my aim.

Consider the following example: On November 5th, 2023, the New York Giants played the Las Vegas Raiders in the National Football League (NFL). During the game, Giants quarterback Daniel Jones tore his ACL while backpedaling to avoid a defensive player attempting to tackle him (event E). Jones was taken out of the game immediately, and the Giants went on to lose the game (outcome O). A question might naturally arise: Would the Giants have won if Jones hadn't torn his ACL? To answer this counterfactual question, we need to look at the closest world in which Jones doesn't tear his ACL. There are several plausible ways to analyze this counterfactual claim:

> ***w_1*:** 5 minutes before the tear, a pass is dropped so that the Giants' offense is not on the field at the time during which Jones tore his ACL. Because he is not on the field, he does not tear his ACL.
>
> ***w_2*:** 2 minutes before the tear, the Giants run a different play that results in a touchdown, getting Jones off the field before he tears his ACL.
>
> ***w_3*:** 2 seconds before the tear, Jones audibles out of the play call requiring him to backpedal. Instead, he hands the ball off to the running back and does not tear his ACL.
>
> ***w_4*:** 0.15 seconds before the tear, the turf miraculously gives way and Jones' leg is able to move deeper into the ground just enough to prevent the tear.
>
> ***w_5*:** At the moment of first fibers in Jones' ACL tearing, the original w_0 is copied and pasted into w_5, where a miracle occurs, simply preventing Jones' ACL from tearing (even with the same weight and force put into it).

Imagine the counterfactual analysis in worlds 1 through 5 all reveal that had Jones' not torn his ACL, the Giants would have won the game. But analyses 1 through 5 are not on a par; they do not equally maximize the similarity between possible worlds. The greater the transitionary period, the less accurate the counterfactual analysis will be. In the image below, we can see that the shaded area representing the differences between worlds shrinks as the transitionary period is reduced until it is finally eliminated in w_5.



In w_1, there are five minutes of different events, unrelated to the question at hand, between the miracle and ACL non-tear. It would be unclear, then, if the differences in these five minutes were causally responsible for Giants winning the game, rather than Jones' ACL not tearing. The same worry exists for worlds 2 through 4, but as the transitionary period gets shorter, the worry of polluting differences between worlds is decreased. The worry diminishes but does not disappear until w_5. Only then have we fully isolated the relevant change.

Moreover, our original worry—that a 'copy; paste' method of counterfactual analysis inevitably raises more issues than it solves—seems unfounded in this case. NFL quarterbacks backpedal on dozens of plays each game without tearing their ACL's. In Jones' case, there would be nothing extraordinary about a player backpedaling exactly as he did without tearing his ACL. No 'clean-up miracles' need be subsequently enacted. Thus, we maximize similarity while respecting Lewis' other desiderata: we "(i) avoid big, widespread, diverse violations of law," (ii) "maximize the spatio-temporal region throughout which perfect match of particular fact prevails," (iii) "avoid even small, localized, simple violations of law," and (iv) "secure approximate similarity of particular fact" (1979, 472).

One might object that I have only argued for a very limited subset of cases—cases in which the event in question (and the alteration of that event) do not include that of which people are *aware*. Perhaps my proposal works in limited cases, but what about other ones—perhaps more common; perhaps more important cases? Consider events involving a decision. It seems like these types of cases require a transitionary period between the miracle and the decision point (the event in question). The thought goes like this: if I am deliberating whether to do X or Y, and I am about to decide on X, miraculously changing my mind to Y at the *moment of the decision* would be odd. That oddity would propagate downstream, creating a radically different subsequent world in irrelevant ways, violating Lewis' desiderata (ii) and (iv). Securing similarity of fact with 'clean-up' miracles would only worsen the analysis, violating (i).

Even if I were to grant this objection, notice just how many cases remain immune. There are many counterfactual questions that can be evaluated without respect to conscious deliberation or decision making (cases like the tearing or non-tearing of an ACL). And since my aim has been to remove the *necessity* of Lewis' transitionary period, these cases alone secure the success of this modest goal.

However, I suspect that even many cases involving conscious decision making can be successfully analyzed without a transitionary period. The experience of making a decision is one of *open-ended deliberation* up until the moment the decision is actually made. So then, to miraculously change the decision made, at the moment of the decision, should not be phenomenologically any differently than how decisions are normally made; doing so should not create the need for downstream 'clean-up miracles.'

Consider a paradigmatic example from Fine: "If Nixon had pressed the button there would have been a nuclear holocaust" (1975, 452). Nixon's pressing the button (or not) comes down to Nixon's *decision* to press the button, so we might reframe the counterfactual statement as 'If Nixon had decided to press the button, there would have been a nuclear holocaust.'

I contend that the 'copy; paste' analysis works without issue in this case and provides a more accurate counterfactual analysis than Lewis' model. The event in question is Nixon's decision to press the button—he may have decided to press the button a second or an hour before actually doing so (this does not matter for our purposes). The relevant outcome is whether a nuclear holocaust would have occurred, had he decided to press the button. We proceed by creating a miracle which changes Nixon's decision *against* pressing the button in our actual world to a decision *for* pressing the button in the possible world.

We have already seen that an analysis without a transitionary period is more accurate because it maximizes similarity leading up to the event in question. What remains to be seen is whether my proposed model creates the necessity of 'clean-up miracles' down the line.

I maintain it does not. The moment of Nixon's deciding not to press the button in w_0 entails that the moment before, he was undecided; both options had to be *genuinely* open to him for it to

be a genuine decision. If we analyze the counterfactual on the 'copy; paste' model, in the nearby possible world w_1, a minor miracle would simply change the moment of Nixon's decision from not pressing to pressing the button. This change in decision, though, would not feel phenomenologically different than deciding to press the button in w_0. Because the prior phenomenology is consistent with both deciding to press and not to press the button, a miracle changing this decision, at the *moment* of the decision, would not create any polluting downstream effects. It appears, then, that the 'copy; paste' model of counterfactual analysis can handle cases involving conscious deliberation as well.

Conclusion

I should admit, however, that there may well be cases that the 'copy; paste' account cannot handle without the creation of many 'clean-up' miracles. In large part, these cases do not interfere with my more general aim: to eliminate the necessity of a transitionary period in Lewis' account of counterfactuals. It seems that, at least in many cases, we may proceed more accurately by examining a nearby world under the 'copy; paste' model. For many cases, our counterfactual analysis may proceed by 'pausing' the actual world at an event, making a copy of that world, miraculously changing the event in question, and then 'pressing play' on both worlds.

References

Byrne, Ruth M. J., (2005). *The Rational Imagination: How People Create Alternatives to Reality*, Cambridge, MA: MIT Press.

Elga, Adam (2000). “Statistical Mechanics and the Asymmetry of Counterfactual Dependence.” *Philosophy of Science* 68 (3):313-324.

Fine, Kit. (1975). Review of Lewis [*Counterfactuals* (Oxford: Blackwell, 1973)]. *Mind* 84: 451-458.

Lewis, David. “Counterfactual Dependence and Time’s Arrow.” *Noûs*, vol. 13, no. 4, 1979, pp. 455–76. *JSTOR*, https://doi.org/10.2307/2215339.

Stalnaker, Robert (1968), "A Theory of Conditionals", in Nicholas Rescher (ed.), *Studies in Logical Theory*. Oxford: Blackwell, 98-112.