Stephen Barker (Nottingham 08)

Abstract: The standard view about counterfactuals is that a counterfactual (  >  ) is true if and only if the  -worlds most similar to the actual world @ are  -worlds. I argue that the worlds-conception of counterfactuals is wrong. I assume that counterfactuals have non-trivial truth-values under physical determinism. I show that the possible-worlds approach cannot explain many embeddings of the form (  > (  >  )), which intuitively are perfectly assertable, and which must be true if the contingent falsity of (  >  ) is to be explained. If (  > (  >  )) has a backtracking reading then the contingent facts that (  >  ) needs to be true in the closest  -worlds are absent. If (  > (  >  )) has a forwardtracking reading, then the laws required by (  >  ) to be true in the closest  -worlds will be absent, because they are violated in those worlds. Solutions like lossy laws or denial of embedding won't work. The only approach to counterfactuals that explains the embedding is a pragmatic metalinguistic approach in which the whole idea that counterfactuals are about a modal reality, be it abstract or concrete, is given up.

Orthodoxy claims that a counterfactual (  >  )[1]

Counterfactuals can be contingently true or contingently false, even assuming physical determinism. Take any counterfactual ( > ) with contingent antecedents and consequences, where ( , ~ ) is physically possible. Such counterfactuals, if true, are contingently true, and if false, contingently false. Furthermore, if contingently false, there are facts that are responsible for the falsity of the counterfactual. We can rightly assert that ( > ) is false because, say, is false. Consequently, we can rightly assert that had been the case, ( > ) would have been true, and thus rightly assert the compound conditional: a of the form ( > ( > )).

Assuming the possible worlds semantics, ( > ( > )) is true just in case the closest -worlds to the actual @ are worlds in which ( > ) is true. That means that the closest P-worlds to @ are worlds such that the closest -worlds to them are -worlds.

The truth of ( > ) at the closest -worlds depends crucially on the facts and laws holding at these -worlds. But just what facts and laws are available there depends on what reading the main counterfactual in ( > ( > )) has, that is, on the similarity metric governing comparative closeness. There are two readings: the reading and the reading. Here is the dilemma. There are many true cases of ( > ( > )), which cannot be certified as true by the possible worlds approach because:

If ( > ( > )) has a backtracking reading then the contingent facts that ( > ) needs in order to be true in the closest -worlds—also present in @—are absent because unconstrained backtracking undermines them.

If ( > ( > )) has a forwardtracking reading, then the laws required by ( > ) to be true in the closest -worlds will be absent, because the closest -worlds cannot have @'s laws and no are available to carry out the work of support.

In short, ( > ( > )) comes out untrue on either alternative. However, the contingent falsity of ( > ) requires its truth. Indeed, ( > ( > )) is perfectly assertible.

In what follows I develop this argument in detail. The paper proceeds as follows. I outline counterfactual possible-worlds semantics in §1 for forwardtrackers and backtrackers. In §2, I present the embedding problem outlined above. I look at solutions that tinker with ideas about laws in §3 and §4. Counterfactual consequent embedding as a general phenomenon is examined in §5. The conclusions of §6 are these: the standard possible worlds approach—which takes the nearest -worlds to be consistent and complete—is unable to solve the embedding problem outlined. The problem cannot be solved by proposing that the nearest -worlds are inconsistent or incomplete. These conclusions hold independently of whether modal realism or erzatism about possible worlds is accepted. I show that other approaches to counterfactuals, call them metalinguistic approaches, do not face this problem. I conclude that counterfactuals are not likely to be about worlds.

According to the possible worlds conception—Lewis (1973)—a counterfactual ( > ) is true if and only if the closest -worlds to the actual world @ are -worlds. We assume, as does Lewis (1973, 1979), that counterfactuals can be evaluated in worlds where physical determinism holds. Counterfactuals come in two basic kinds: and . Take the counterfactual (1) below evaluated a little after 12 pm with respect to a window, ninety storeys above a concrete road, with no means of breaking a fall:

> (1) If I had jumped out this window a moment ago, I would have died.

(1) has a reading on which it is clearly true. Intuitively, it is that reading in which we envision ourselves as simply leaping out the window moments before and falling rapidly to the ground as a result. Consider (2) evaluated under the same circumstances:

> (2) If I had jumped out this window a moment ago, it would only be because I had first learnt how to base jump, etc, and so I would not have died.

(2) also looks assertable. In (2), we are envisioning the most likely way that our leaping out the window could have come about. (1) and (2) may look incompatible with each other, but they are not really. Each features a distinct reading of the counterfactual locution. (1) has a forwardtracking reading, (2) a backtracking reading. The possible-worlds approach explains this distinction in terms of a difference in the comparative similarity relations being invoked

in each case. The comparative similarity relation governing (1) guarantees that the nearest antecedent worlds diverge a little before 12, when some change in my brain leads to my leaping out the window. The comparative similarity relation governing (2) guarantees that the nearest -worlds diverge from a much earlier time, say, a year ago, when I start training with parachutes with the aim of leaping out the window.

The possible-worlds conception needs some general set of respects of similarity that will deliver late divergence in the case of (1) and possibly early divergence in the case of (2). Lewis (1979) provides the generally accepted analysis of the forwardtracking reading. Lewis proposes a set of weighted respects of comparative similarity whose most important facet is a trade off between maximising regions of perfect match of particular fact and law violations. Applied to deterministic situations, the result is that the closest -worlds are ones sharing history with @, exactly, up to a point not long before t , the time of , at which point a minor miracle occurs, and a divergence to is activated. Thereafter, no other miracles occur in the closest -worlds.

In all this, is a term of art. The laws violated by the various miracles in -worlds are not laws of the -worlds. They are @'s laws: the miracles are miracles, law violations, from the point of view of @. The standard view is that laws cannot be violated and remain laws. This is certainly the case for Lewis (1973) who accepts the best-system analysis of laws according to which law-statements are those universal quantifications that feature in the best, most economic and explanatorily rich description of a world. The laws of -worlds, where @'s laws are violated must be different laws. What are these laws? Lewis does not say. And for the purposes of evaluating simple counterfactuals, there is no reason to know.

We shall assume Lewis's analysis of forwardtrackers in what follows. The embedding argument outlined in §0, however, does not depend on any particular detail of it. The crucial fact is that forwardtrackers, being late divergence conditionals, must, assuming determinism, contain miracles relative to the actual world. That's the key fact.

How do backtrackers work? Lewis (1979) does not tell us, nor does anyone else. So let's quickly develop a theory. An initial analysis of the backtracker comparative similarity relation emerges from the thought that backtrackers involve an extrapolation backward from the antecedent time that changes facts but keeps @'s laws. Take the temporal slice of the actual world @ at the time t , of the antecedent . The idea then is that the closest backtracker -worlds will be worlds that have a temporal slice at t that differs minimally from @'s slice at t and preserve @'s laws. Or in other words:

> ***B1***: The closest backtracking -worlds to @ are worlds with a minimal modification of the microstates of @ at t so as to instate , which preserve @'s laws before and after t .

***B1*** is meant to explain the truth values assigned to backtrackers, so should explain the truth of (2), if it is right. But it doesn't do that. Given the minimal modification of the microstate allowing me to jump out the window at 12, the lawful extrapolation backwards does not have me setting up safety systems, and so the lawful extrapolation forwards has me breaking my neck seconds later. So, given ***B1***, (2) is false.

Here is a more promising proposal than ***B1***. Call the the kind of development of events that leads to in the nearest -worlds. In the case of (1), the scenario is just me suddenly leaping out the window. In the case of (2), the scenario begins much further back in time with my learning to base jump and then leaping on the day with a parachute. Both of these developments, as divergences from the actual world @, have zero probability, since they are excluded by the laws and facts at any prior time. But there is a difference. At a certain level of detail of description, the first scenario is vastly more improbable than the second. The forwardtracking scenario just has me getting up and leaping, a very unlikely thing for a sane human in ordinary circumstances to do. The second scenario does not represent a particularly improbable event—sane individuals base jump all the time. This judgement of comparative likelihood can only occur from a level of description that does not open itself to complete detail about physical realisation of the causal paths. Still it can be made. Furthermore, we can see ask the following question: on the assumption that I had jumped, what is, in this sense, the most likely way that could have come about? The answer is the second scenario. This is what (2) is expressing. Here then is the proposal about backtracking comparative similarity that works:

***B2***: The closest backtracking -worlds are the ones that allow the most probable scenario of causal development to from any prior point in @ given some level of description; minimize miracles.

Because backtrackers are about causal developments and their degrees of likelihood they presuppose causal knowledge. This is fine from the point of view of the counterfactual theorist of causation. It's forwardtrackers, not forwardtrackers, that ground causation.

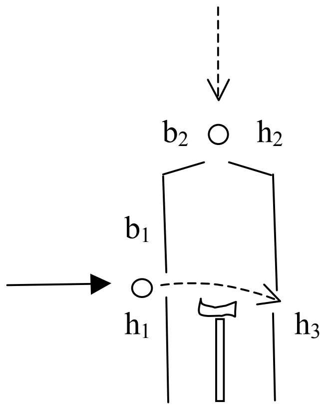
In asserting a backtracker, ( > ), the scenarios of development leading to that are probable, are only likely relative to a degree of detail. If you pursue the question of how exactly comes about then, assuming determinism, you can always undermine a backtracker unless the divergence is at the very beginning of the universe. Suppose a ball has fallen from a great distance and lands in a certain spot. With the prospect of physical determinism firmly in our minds, we might assert:

(3) If had not fallen, all prior history would have to have been different.

A philosophical backtracker like (3) is one in which the envisaged scenario of development can withstand any question about the causal path at any level of detail. At the highest level of detail, what is the most likely path leading to 's not falling? The answer is some change in the first instant of the universe, that is, a change in what we might call the initial conditions. We know that any change here will almost certainly produce global consequences later. Philosophical backtrackers are then not very informative; they are just ways of conveying that the universe is causally deterministic.

We now have some idea of the forwardtracking and backtracking distinction. That is enough to set up the problem outlined in §0. In what follows, I describe circumstances in which a counterfactual of the form ( > ) is contingently false, and would have been true if had been the case. That is, a counterfactual of the form ( > ( > )) is true. But, for the reasons given in §0, the possible-worlds approach is unable to certify its truth.

Consider the following set up—see the diagram below. A metal cylinder has two side-holes $h_1$, $h_3$, and a top hole, $h_2$. By the side hole $h_1$, a ball $b_1$ is supported. If $b_1$ is propelled with force , it will, given the gravitational field and air density, pass through $h_1$ into the cylinder, passing through $h_3$ on the other side and out of the cylinder. The experiment is done a multitude of times and each time $b_1$, propelled with force , goes through the holes. Suppose now that there is another ball, that has been approaching from some vast distance, and which has recently been caught in the earth's gravitational field. It falls toward earth and drops through the top hole $h_2$ of the cylinder and lodges itself at a point that is in the path that $b_1$ takes if propelled, on a pillar with a ball-catching pillow on top of it. Here is the image:




At time , the instant after $b_2$ lands and blocks the path for $b_1$, the counterfactual (4) is false:

(4) If $b_1$ had been propelled with force , it would have passed through $h_3$.

(4) is false, but contingently so. Certainly, there are physically possible worlds in which (4) is true, but we can say something stronger than this. We know why it is false. It is false because $b_1$'s path was blocked. Its path was blocked because $b_2$ dropped through $h_2$ after its fall over a great distance. So, (4) is false, because $b_2$ fell. That particular because-statement seems to commit us to a counterfactual:

(5) If $b_2$ had not dropped, (4) would have been true.

The truth of (5) is just part of our articulation of the contingent falsity of (4). In other words, we specify circumstances such that, had they been the case, (4) would have been true.

Given the truth of (5), we are committed to the truth of (6):

(6) If $b_2$ had not dropped, then, if $b_1$ had been propelled with force , it would have passed through $h_3$.

The transition from (5) to (6) is unexceptionable. (5)'s truth demands that we evaluate a counterfactual (4) relative to antecedent-worlds.

That's the set up, now for the problem. (6) is a compound, consequent-embedding counterfactual—of the form ( > ( > )). We can take it that the embedded conditional (4) has a forwardtracking reading. That's because conditionals like (4), when we take them to be true, are just the standard forwardtracker conditionals. That leaves us with the status of the main conditional, (6) as a whole. Its reading will determine the character of the nearest ( )-worlds. (6) is fairly obviously a forwardtracker; we are not concerned with the most likely way that $b_2$'s not falling comes about but with the consequences of $b_2$'s not falling. But let us not assume that in the argument that follows. We consider the two available hypotheses: first that it is a backtracker, second that it is a forwardtracker. I show that on both readings, assuming possible worlds semantics, (6) cannot be certified as true, for the reasons given in §0 ). On the backtracking reading the factual basis, in the ( )-worlds for the embedded conditional are undermined. On the forwardtracking reading, the nomic basis for the embedded conditional are undermined in the ( )-worlds.

Assume that the main conditional in (6) is a backtracker. Backtrackers are about the most likely scenarios of development to the antecedent condition, given a level of detail about scenarios of development. The only way to get any assurance of no violation of law in the nearest antecedent worlds is to treat (6) as we did the philosophical backtracker (3). That means that the nearest ( )-worlds are ones where facts all the way back to the first instants of the physical universe are changed so as to preserve physical law. So a change very close to the birth of the physical universe is required. The early universe is (relative to later stages) very simple and highly compressed. Any change here will concern the whole universe at that stage, and have consequences spread out across the whole universe at each later stage. The existence of $b_1$, its falling, the experimental set up of the cylinder, the humans who created the cylinder, and so on, all have a common cause in the very simple state of the first instances of the universe. None of these things exist in the early universe, not even the particles that make up these things. Any change at this stage will generate a globally different universe.

If that's right, then on a backtracking reading, (6) is false. If $b_2$ had not dropped, would (4) have been true? If $b_2$ had not dropped, and the whole prior history of the universe was different, and $b_1$ would almost certainly not have existed, no experimental set up, no humans, and so on. Under those circumstances can we straightforwardly certify the truth of (4): if $b_1$ had been propelled, it would have passed through hole $h_3$? Since there is no set up, no $b_1$, we can almost certainly say it is false. We have completely undermined the factual basis for the embedded counterfactual (4). So, in a backtracking reading the truth of (6) cannot be certified.[3]

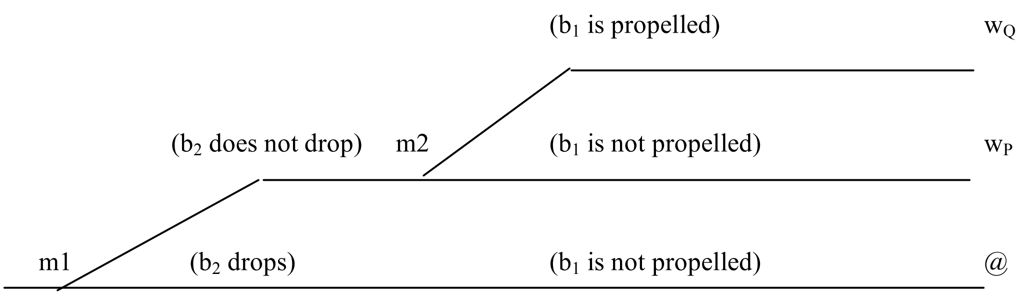
Let us then conclude that the main conditional in (6) cannot be a backtracker. So it must be a forwardtracker. That means that the nearest ( )-worlds are worlds that diverge relevantly recently before the passing of $b_2$ through the hole. The relatively late divergence has this desirable feature. We preserve the contingent facts of the set up that we need for the otherworldly truth of (4). In particular we preserve the fact that $b_1$ exists, that there is a cylinder present, that the holes are in position, that there is a certain gravitational field, and so on. We also preserve the fact that $b_1$ is not propelled. There is no (recent) causal relevance between $b_2$'s dropping or not dropping and $b_1$'s being propelled.

However, the price we pay for preserving these facts, lost on the backtracking reading, is that the laws of the actual world @ are violated. Since miracles are impossible—we are currently assuming that strict laws have no exceptions—it follows that the laws in the ( )-worlds cannot be those of @.

[3] Could we not demand of the back-tracking reading require that particular fact be kept as close as possible to the events in @? We could, but that will not help. It will very likely be physically impossible to preserve law, knock out the event of b2's falling and keep all the other local facts relevant to the set up.

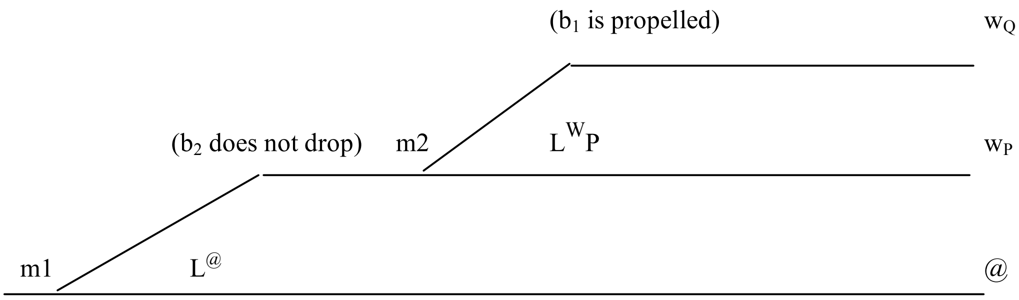
Our concern now is what laws get violated to secure late divergence to $b_2$'s not dropping? Here are some miracles that might do the trick. One is that the gravitational field spontaneously warps sending $b_2$ off to the side. Another is that the air molecules on one side suddenly become more massive causing $b_2$'s divergence. And so on. There is no particular reason to suppose that violating one law rather than another will secure a closer ( )-world. Some violation m1 will have to occur.

The counterfactual (6) also requires that we go to the closest ( )-worlds relative to the ( )-worlds. These worlds will be late divergence worlds branching from the ( )-worlds, as in:




There will be a miracle m2 involving violation of the laws at the ( )-worlds. Apart from this violation, the ( )-worlds involve no other violation of the laws of the ( )-worlds. From m2 on, the ( )-worlds develop precisely as if they were being governed by the laws of ( )-worlds. So to know if (4) is true, we have to know that the laws in the ( )-worlds will do the job of supporting (4). Here is a suspicion that they cannot. The laws that are violated in getting us to the ( )-worlds are amongst the laws that are relevant to the dynamics of $b_1$ in the ( )-worlds. For example, gravitational laws may be violated in order for $b_2$ to diverge sideways in its fall. But gravitational laws are also relevant to $b_1$'s motion. It may be a conservation of energy laws that are violated in the divergence to ( )-worlds, but these laws are also relevant to the motion of $b_1$.

Schematically, we can sum up the situation in relation to laws thus:




The law of the actual world @, $L^{@}$, is violated in $w_P$. $L^{W}p$ is the counterpart law, as we might call it, of $L^{@}$ in $w_P$. The miracle, m2, is a violation of laws in $w_P$, but we need not consider m2 here. The development in $W_Q$, in particular the progress of $b_1$ in its path of motion through $h_3$, is determined by $L^{W}P$. It is crucial to know exactly what $L^{W}P$'s character is if we are to be assured that $b_1$ reproduces its actual world motion.

What we require is a general principle that provides us with a function from $L^{@}$ to its counterpart $L^{W}P$. If there is no independently specifiable function, there is no objective fact about whether or not (4) is true in the nearest ( )-worlds. I argue now that there is no function that secures the result of preserving $b_1$'s motion. The violation of $L^{@}$ undermines the truth of ( > ) is the closest -worlds.

The idea that springs to mind is that counterpart laws of the nearest -worlds, $L^{W}P$, will be minimal variants of the actual world laws $L^{@}$. The laws in the ( )-worlds will only differ as much as is necessary in order to allow the miracles of divergence to occur. Let us assume that the law violated to get to the ( )-worlds is the gravitational law. Let us assume a Newtonian world, so the law determining magnitude of force is:

and the direction of the gravitational force vector is determined by the centres of gravity of the attracting masses. We shall assume that it is this directional vector law that is violated. What then is the minimal variation of the directional-vector law ? The following, M is as minimal as we can get—in what follows is the space-time region at which the violation of occurs:

> M: The direction of force on two masses $m_1$ and $m_2$ is determined by the line between their gravitational centres, unless $m_1$ or $m_2$ is the mass of an object at , in which case the direction is degrees.)

Here, the value is whatever value we want the force to be to get $b_2$ in the right position. The idea now is that this is the law governing the direction of gravitational force in ( )-worlds. Furthermore, in allowing this law to drive development—along with other dynamical laws—in the closest ( )-worlds, we get the result that $b_1$ goes through the hole, as counterfactually predicted, since the force vector will have the expected magnitude and direction. The compound (6) will come out certifiably true, as desired.

Let's put this in more general terms. A physical law is paradigmatically a function. A law, qua function, L( , ), links sets of quantities to a quantity . So if L( , ) is Newtons's gravitational force law, is force , and is the set comprises masses, $m_1$, $m_2$, and distance . The determinants are particular forces, particular masses and distances. We can think of these as properties. stands for the particular functional relation holding between determinants ′ onto ′. Suppose that L( , ) is the law violated in the closest - worlds. The violation is at the space-time region by object $b_2$. Object $b_2$ has some determinant property but not the corresponding , but another determinant . The corresponding minimal law variation of L( , ) is the law whose function matches perfectly that of L( , ) except for the case of ′ instantiated at , for which the corresponding value is . Call this law $L^M$. At the nearest -worlds the violation we need gets us an exception to L( , ), but there are no more exceptions to L( , ). Furthermore, in the closest -worlds, object motion is guided by $L^M$ from afar—it may not itself be a law of the closest -worlds—and will mimic L( , ). In which case, ( > ( > )) comes out true, as desired.

That's the program for the minimal variation strategy, but it won't work because generalities of the form $L^M$ cannot be laws. One might contend that that if we are Humeans there is no barrier to deeming functions like $L^M$ laws. They are regularities in some sense. Why can't they be members of a best system for describing a world? But the idea that any kind of regularity can be a law does not wash with the basic explanatory norms of physical science. Let us see how that is so.

In relation to L( , ), we can ask, why doesn't object $b_2$ at , which has determinant property not have determinant ? There are two replies, but on both we undermine the idea that $L^M$ is really a law. The first reply is that there is no reason why $b_2$, with property ′ lacks ′. If so, it's physically accidental that $b_2$ lacks ′. In other words, we have just admitted that a nomic regularity has an exception. So $L^M$ really is not a law.

The alternative answer is that there is a reason why $b_2$ does not instantiate ′. The reason is that it's at . But then what is it about the position that makes it physically significant, that is, causally relevant to possession of determinants of ? One answer is that it is just a brute feature of . But then we are assigning physical significance to bare particularity itself. That's objectionable. In physics we do not attribute nomic relevance to bare particularity: physical objects have the status of bundles of generic properties. Physical duplicates identically related to other physical duplicates have the same nomological features. Our law $L^M$ just looks like a law that imputes significance to haecceities or bare particulars: to the particular . But to impute such significance is really, from the physical point of view, just to allow exceptions.

Some physical particulars physically significant. For example, the singularity at the beginning of the universe has huge physical significance. But the singularity, , has significance in virtue of its physical properties. A theory, General Relativity, given certain empirical facts, can be used to predict that exists. Our current physically significant particular, , is not nomically predicted. It's just brutely significant.

If we are really going to find a physical law that is a minimal variant of L( , ) we need a variant that will explain why being at had this affect on the physical forces governing $b_2$'s motion. What we need is a physically respectable that has a determinant property ′ that possesses, such that if objects have relations to space-time regions to determinants ′ of , their behaviour is changed. cannot be one of the existing physical quantities. In order for a determinant ′ of an existing quantity to be instantiated at , we need to introduce violations to get these determinants into position. But that would just involve other law violations beyond that which is the change of direction of $b_2$.

So it seems has to be an alien quantity. That means that in the nearest ( )-worlds, a new fundamental quantity appears. No physical quantity can exist without laws governing it. (If is not governed by laws, then our postulation of it amounts to no more than the imposition of brute physical significance to particulars in the sense we have already rejected.) There must be laws governing and laws governing its relation to other quantities, such as mass and gravitational force. That's the we need, but there is serious doubt that the idea of such a physical quantity is coherent.

has only one of its determinants, ′, instantiated, only once, at a small region , very late in the universe. No real fundamental physical quantity could have that miniscule foothold in the world. The configuration of forces at produce ′; these conditions never appeared earlier or later, nor were comparable conditions produced at any other time at which other determinants ′′ of appeared. If there really are laws governing in relation to other forces, will have a range of determinants, ′, ′′, ′′′, etc, which will be produced by variants of the physical forces at . The conditions at would have slight variants, if not perfect duplicates, elsewhere. All you need is an object with a certain mass under a gravitational force. Why haven't these slight variants of produced determinants of ? Why, for example, didn't the other balls that fell slightly before $b_2$ under almost identical physical forces, give rise to determinants of ? cannot be a fundamental physical quantity interacting with others in the world if its interactions are virtually infinitesimal.[4]

The minimal variant approach fails. It fails because a world whose property distribution conforms perfectly to @'s laws, except for some isolated cases, is not going to be a world fully governed by distinct laws.

---

[4] There is little here to constrain the laws that supposedly govern . No Humean about laws could countenance this. A non-Humean would have to say a vast set of nomic relations between uninstantiated determinants of exist. But what ties any of these down to concrete reality?

If the minimal variant approach to laws fails, where can we go from here? We have been assuming that laws cannot have exceptions. Maybe the way out of the embedding problem is to countenance laws with exceptions. Thus the law L( , ) that is violated to usher in divergence to ( )-worlds lives on in those worlds with an exception but still ensconced as a law in those worlds. Humeanism about laws is typically taken to imply strictness. Some philosophers deny strictness. Braddon-Mitchell (2001) is a case. According to Braddon-Mitchell, the best system analysis of law does not have to be taken with the requirement that generalities in the best system be true. By including L( , ) we capture more truth economically even though it predicts falsely, in some cases, what happens. So exceptions are allowed. What is required of the best system is that with the minimum of empirical fact, most of the rest of the worldly facts are recoverable. The trick is optimizing the balance between minimizing falsity and maximizing recovery of fact.

The main problem with lossy laws is that a world governed by them cannot be a physically deterministic world, and our present concern is counterfactuals evaluated in physically deterministic worlds. Lossy laws cannot be deterministic since it is not consistent with the lawful status of a deterministic law, L( , ), that a determinant ′ is instantiated by an object , but lacks a determinant ′. But that's exactly what lossy laws allow. A physically deterministic world cannot be a world governed by lossy laws. A lossy-law world is a non-deterministic world. This is not to say that it is a physically indeterministic world in the sense that it has objective chances between zero and one. A world of chance, as proposed by quantum mechanics, is a world of nomically grounded probabilities. The wave function is deterministic. Probabilities can be derived, determinately, from states of the wave function. But no such probabilities can be derived from lossy laws. A world of lossy laws is a world lacking both physical necessitation and objective probability.

Why not contend that the actual world, @, is deterministic, and that the nearest -worlds are governed by lossy laws? So, the same laws can hold in the nearest -worlds as in @, they just hold as lossy laws in the -worlds, whereas they hold as deterministic laws in

@. According to this proposal, we conceive of the world as physically determined, but in contemplating counterfactual hypotheses, we shift to an a-deterministic conception of the world. But is the deterministic character of the world so fragile? When we think of a world as deterministic, it remains so under what are trivial counterfactual suppositions. After all, it is , when we suppose that $b_2$ had not dropped. A global shift in character, the loss of physical determinism, does not ensue just from this slightest of suppositions. If a world is deterministic, it remains so under a range of (intrinsically law-abiding) suppositions.

The otherworldly lossy law proposal also faces the issue of whether lossy laws can support counterfactuals. Had been the case, would ( > ) have been the case? The nearest -worlds are worlds with lossy laws. We may doubt that a world lacking physical necessitation is a world in which would-counterfactuals can be true. Suppose you are told that L( , ) is a lossy law in the nearest -worlds. It is not physically necessitated that a given object that has determinate property ′ has determinant ′. If nothing in fact requires that the lossy law is fulfilled, then we should not expect that things would be any different under a counterfactual supposition that another object * is ′. So the counterfactual below is not true in the closest P-worlds:

′

Believing in his counterfactual requires believing that were * to be ′, the lossy law, L( , ), would cease to function as a lossy law for the new counterfactual case, that is, it would become a strict law for this case. But laws do not change their character because of a law-abiding counterfactual supposition. If that is right, then if the closest ( )-worlds are lossy-law worlds, we should have no reason to think (4) is true in those worlds.

That's the dilemma argument completed. I have used Lewis's conception of forwardtracking in terms of miracle minimisation of perfect match maximisation, but the argument did not depend on the details of that account. It just depends this thesis: assuming determinism and that counterfactuals are about what goes on in antecedent worlds, the antecedent worlds of forwardtrackers must contain violations of @'s laws.

At this stage, the friend of possible worlds semantics may seek to rule out compound counterfactuals, the cause of all their woe. Perhaps, the offending counterfactual (6) and its kin, are misleading as to their real logical form. They are not as they seem—compound counterfactuals—but something else. The salient idea is that they are really counterfactuals with conjunctive antecedents. We embrace :

> : Counterfactuals of the surface form ( > ( > )) really have logical form (( & ) > ).

So (6) really has the form of (7):

> (7) If $b_2$ had not dropped and $b_1$ had been propelled, then $b_1$ would have passed through the hole $h_3$.

(7) presents possible-worlds semantics with no difficulty at all.

How is this response squared with the commitment of contingency of counterfactuals? It might be claimed that all that's asserted in claiming that ( > ) is contingently false is that ( > ) is false, but it might have been true. And all that's asserted herewith is: . None of these claims, one might contend, commit us to a counterfactual ( > ( > )). Similarly, we might argue that the true -statement, , does not imply (6).

This response is seriously disputable. Let us tackle first the point about contingency. It would be odd to claim that (4) is false, as a matter of contingent fact, but not to be able to specify circumstances, , such that, had been the case (4) would have been true. If we think of counterfactual propositions as made true be worldly circumstances, and we do, and we recognize that some of the circumstances required for the truth of a given counterfactual are absent, then it would appear to be just part of the idea that a counterfactual has truth-conditions, that we can assert that had the non-obtaining circumstances been the case, the counterfactual would have been true. The claim that fails in the case of (4) is highly implausible.

The claim about the -statement, such as, , never imply counterfactuals is also highly disputable. We know why the counterfactual (4) is false. We can point precisely to one of the conditions that is responsible for its falsity: $b_2$ falls. We also know that $b_2$'s dropping did not undermine any other potential ground for (4)'s falsity. It's not the case that $b_2$'s falling, given other facts, ensured (4)'s falsity and also ensured that other sufficient conditions for (4)'s falsity did not occur. In other words, $b_2$'s dropping did not undermine, or pre-empt, other grounds for (4)'s falsity. Add to these facts the following eminently plausible principle:

> If a -statement of the form is true, and 's being the case has not pre-empted any other ground for , then the counterfactual ( > ) is true.

The principle applied in the case of causation implies that where causes , and does not pre-empt any other cause of , then had not occurred, would not have occurred. That looks like a highly compelling principle—indeed, it has the status of a methodological principle guiding thought about counterfactuals and causation. Applied to the facts of our experimental set up, it implies that had $b_2$ not dropped, (4) would have been true. So it implies that (6) is true.

We have reinforced the idea that the contingent falsity of (4), and the fact that (4) is false because $b_2$ dropped imply that (6) is true and is really a compound embedding. The translation schema cannot then be applied to (6). We can add to these considerations a brief list of other phenomena that indicate strongly that conjunctive paraphrasing won't work generally.

There are many sentence types for which counterfactual analyses are appropriate. These sentences can be embedded in the consequents of counterfactuals. That implies consequent embedding. Here is a simple case. Dowe (2001) argues that statements of prevention, are counterfactually analysed as

It is acceptable to assert sentences of the form:

—abbreviated:

. That commits one to the following embedding:

.

By simple counterfactual logic, we can derive a consequent embedding again.[5]

Compound counterfactuals like the following resist conjunctive paraphrasing: assume that Juan and Juanita are physical duplicates and always do the same. Suppose that the setup is that one of the two must be chosen to perform a task. The task is to press either button A or button B. It seems we can assert:

But the conjunctive-antecedent paraphrase of this conditional in nonsense:

.

I conclude that the conjunctive paraphrasing of consequent embeddings proposed in won't work.

It would appear that the possible worlds approach to counterfactuals is simply inconsistent with a commitment to physical determinism and consequent embeddings, and that we need the latter to explain the contingent falsity of some counterfactuals. I conclude that there is good evidence that counterfactuals are not about possible worlds.

So what approach to counterfactuals can deal with embedding? To answer that question we need to sum up what we have learned. Call a sequence of suppositions a series of nested suppositions: supposition of $_0$, then, within its scope, supposition of $_1$, then within its scope supposition of $_3$, and so on. A sequence of suppositions grounded in the actual world @, are just ones undertaken at that world. Law-abiding suppositions are suppositions

---

[5] Those attracted to a counterfactual analysis of causation (Lewis 1973a) will face a similar problem with embeddings like: .

that, in themselves, do not violate any law. The laws that are accessible to a supposition are the laws that can be appealed to in deriving conclusion about what would happen given the supposition. I suggest, what we have learned is the thesis ( ):

: In any sequence of law-abiding suppositions grounded at @, the laws that are accessible at any stage are simply those of @.

Thus, laws are preserved through (law-abiding) suppositions. The possible worlds approach to counterfactuals implies that is false. But we have seen this wreaks havoc in the case of compound conditionals. The challenge for any theory of counterfactuals is to explain how holds. This paper ends with a brief discussion of the possibilities.

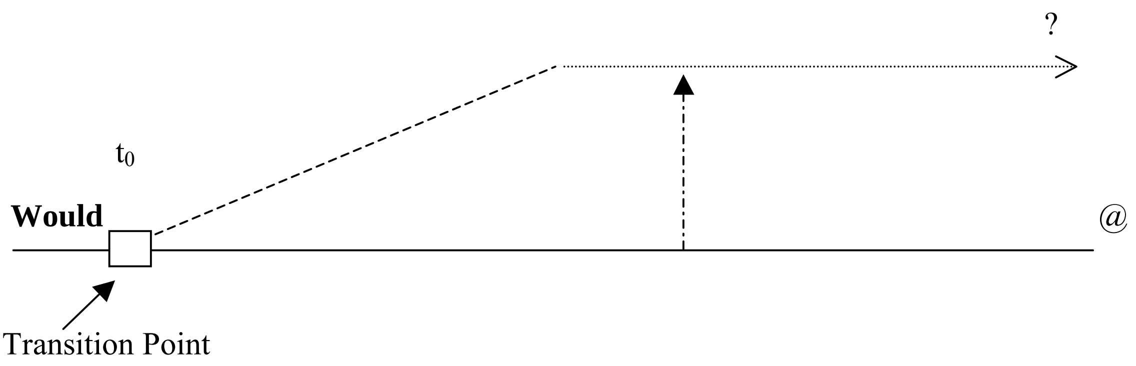
Those who love world-semantics may want to open the door to an worlds-semantics. The nearest -worlds are inconsistent. They are worlds which preserve the laws of @, but also contradict those laws. There is no joy here. In our example, the nearest ( )-worlds are ones in which $b_2$, contrary to law, must diverge from its path of fall. But if the law is not to be violated, it must also retain its path. That would require that $b_2$ had two paths, the one it actually takes and the alternate path. But this would mean that the nearest -worlds would have to contain all the truths of the actual world and more. But that is absurd. We cannot assert that had $b_2$ not fallen through the top hole $h_2$, it would still have fallen through $h_2$. But this counterfactual would be certified by an impossible-worlds semantics.

If the nearest -worlds are not inconsistent, perhaps they are incomplete. One might hope that regions of indeterminacy in a world could allow a divergence to ( )-worlds without law violation. That is not so. In general, the nearest -worlds to @ match the actual world prior to some point $t_0$. The nearest -worlds then are factually determinate up to $t_0$, thereafter pockets of factual indeterminacy may arise. But a physically deterministic world, with complete factual determinacy up to some temporal point, and factual indeterminacy thereafter must contain law violation. The violation will have this form: there is a law, L( ,

) with some object such that has a determinant property ′ but there is no fact that has ′ and no fact that lacks ′. That is sufficient for L( , ) to be violated.

A worlds-semantics doesn't work. That conclusion is independent of modal realist or eraztist doctrines. The latter are about the metaphysical status of worlds. The problems we have uncovered pertain to the logical and physical structure of worlds. Let us cast worlds aside and look elsewhere. The problem of evaluating counterfactuals under determinism, given a forwardtracking reading, is dealing with the inconsistency of law and fact that is inherent in divergence. The possible worlds approach defines truth-conditions for counterfactuals in terms of conditions in which the inconsistency is eradicated. The alternative I now explore does not adjust the inconsistency embodied in divergence, but instead views counterfactuals as incomplete representations of divergence, representations that never register the inconsistency. The inconsistency is not registered, because counterfactuals represent a divergence only up to a certain degree of detail. Counterfactual-judgement makers are typically not interested in sharpening the representation to reveal the contradiction. This idea can be made to work within a so-called or approach.

The basic idea is that counterfactuals express judgments that are extrapolative: using laws, information about facts at a temporal point $t_0$, prior to the antecedent time, t , a representation of worldly development forward is constructed, which incorporates, a supposition that reality from $t_0$ to , and, any facts if those facts are with reality's developing from $t_0$ to :[6]




[6] The approach is drawn from Barker (1999), which is based on Kvart (1986) and Dudman (1994).

The diagram is not the representation of possible worlds but the structure of a judgement with relevant informational factors. The position of $t_0$ is fixed contextually. On a backtracking reading $t_0$ is slid a long way back, on a forwardtracking reading it is slid closer. Barker (1999) develops an algorithm for determining conditions under which  is cotenable, that is, the conditions under which it can incorporated in the judgement.

The divergence from reality at $t_0$ is conceived of as a natural causal sequence, at a certain level of detail: name it . If the development spanning $t_0$, in the region represented by the blank box, were specified in full physical detail it would be inconsistent with law. But that fact of contradiction is simply not registered in the construction of the judgement in normal contexts. Speakers have no interest in doing so.

Broadly speaking, judgements of counterfactuals work in these terms: U judges ( > ) true iff U can find a law-based implication, ( & → would ) such that U judges: is true; relative to the degree of detail given in the scenario, is cotenable with . A game can ensue between assertor and assessor depending on how much detail they want to invest in . We may suppose that conversational interest fixes a level of detail. In a known deterministic context, an uncooperative audience can always raise the level by pursuing questions about the exact causal path in , which inevitably leads to a collapse in the defence of the forwardtracking judgement.[7]

The approach has no problem with compound counterfactuals. Embedding of (non-counterlegal) suppositions preserves laws. above is validated. U judges ( > ( > )) true if U can find a law-based implication, of the form with true and , and judges that the cotenability conditions hold, where - and with ′ are the scenarios of development:

( & → ( & ( & → )))

is cotenable with , is cotenable with ′ .

Applied to our test case, we judge (6) true if we can find a law-based implication:

( ) & → ( & (( ) & → (( )))

---

[7] Compare here Lewis's (1999) contextualist conception of knowledge: The sceptic creates a context in which knowledge disappears by raising the threshold governing relevant alternatives.

In this case, includes facts about the experimental set up. is a subset of these facts. The cotenability condition, in the simplest case, is satisfied if no part of the scenarios that record the transition of reality to ( ) or ( ) undermines the facts and . That makes judgement of (6)'s truth relatively unproblematic, as it should be.

The worlds idea of counterfactuals was a beautiful idea. But we have good evidence to believe that the program cannot be carried out. Counterfactuals under determinism are partial representations of an inconsistent reality, not complete representations of reality whose inconsistency is somehow adjusted away. The possible worlds program is committed to the latter idea. A pragmatic, metalinguistic program can enable us to articulate the first idea. It will deliver the preservation of laws under (non-counter-legal) suppositions, and that's what we need to understand embedding and the contingent truth and falsity of counterfactuals.[8]


Barker, S., 1999. 'Counterfactuals, Probabilistic Counterfactuals, and Causation', 108: 427-469.

Braddon-Mitchell, D. 2001. 'Lossy Laws'. 35: 260-277.

Dudman, V. H. 1994. 'On Conditionals'. *Journal of Philosophy* 91: 113-128.

Dowe, P., 2001: 'A Counterfactual Theory of Prevention and 'Causation' by Omission', 79: 216-226.

Kvart, I. 1986. . Indiannapolis: Hacket Publishing Co.

Lange, M. 2005. 'Laws and Their Stability,' : 415-432

Lewis, D. K., 1973. . Cambridge Massachusetts: Harvard University Press.

___________1973a. 'Causation' (plus postscripts), in Lewis 1986: 159-240. Originally published in 1973 in 70.

___________1979. 'Counterfactual dependence and Time's Arrow' (plus postscripts), in Lewis 1986: 32-66. Originally published in 1979 in 13.


[8] The conclusions of this paper dovetail nicely with Lange's (2005) idea that physical laws are identified by their robustness in relation to non-counterlegal counterfactual suppositions.

___________1986. . Oxford: Oxford University Press.

___________1999. 'Elusive Knowledge', in ,
Cambridge: Cambridge University Press: 418-45.