PHIL-D-22-00480

A New Probabilistic Account of Counterfactuals

Counterfactuals and subjunctive conditionals are enormously important in philosophy, especially in philosophy of science. They are crucial to accounts of laws, causation, dispositions, probability, measurement, knowledge, reliability, decision, the direction of time and much else. In view of this it is surprising that it wasn’t until mid-20th century that serious proposals for the logic, semantics, and analyses of counterfactuals began to appear. The subsequent literature is vast with many proposals, counter examples to proposals, and replies. The most prominent account is David Lewis’ analysis of counterfactuals in terms of possible world similarity. This paper focuses on that account and the problems that it and related approaches based on similarity among possible worlds encounter and offers a new theory that solves those problems by replacing similarity with objective probability. The new account bases counterfactuals on statistical mechanics and thus answers Lewis’ question of how the temporal asymmetry of counterfactuals is related to the second law of thermodynamics. By doing this it accounts for the temporal asymmetry of causal counterfactuals and shows how counterfactuals explain why we can influence the future but not the past and why we can know much more about the past than about the future. The account construes statistical mechanical probabilities in terms of Lewis’ BSA of objective probabilities and laws and thus integrates counterfactuals into that account. By doing this it shows how counterfactuals together with laws and objective probabilities supervene on the distribution of fundamental properties in space-time (the Humean Mosaic).

Before getting into accounts of counterfactuals I want to note that we will heed Alan Hajek’s advice to distinguish two different projects for a theory of counterfactuals.[[1]](#footnote-2) One project is producing a theory that accounts for the counterfactuals that English or other natural language speakers find assertable and acceptable. The data for such a theory include speakers’ intuitions and utterances and examines various languages. This may involve describing the various ways counterfactuals are expressed, and different kinds of counterfactuals and other subjunctive conditionals. It also may involve examining how context plays a role in determining the truth conditions of counterfactuals. This difficult project is the business of philosophy of language and linguistics.

A second, although not completely distinct and not incompatible project, is to produce an account of the kind of counterfactual that plays a central role in philosophy of science. These are connected with laws, causation, influence, measurement, time’s arrows and so on. Call his kind ‘the core counterfactual.” The core counterfactual may not correspond exactly to any natural language conditional, but it must be close enough so that the relations between the two accounts can be made clear. Ideally, it will be possible to account for the natural language counterfactual in terms of the core counterfactual together with pragmatic and psychological factors. But as Hajek says these two projects can pull theorizing in opposite directions. For the most part it is the second project that for the engages Lewis and is the focus of this paper. But I will also keep an eye on the first project.

Nelson Goodman in *Fact, Fiction, and Forecast* provided one of the first proposals of an account of counterfactuals.[[2]](#footnote-3) Goodman claimed that a counterfactual with a false antecedent like “If the match had been scratched it would have lit” is true only if the antecedent together with certain truths and laws imply the consequent. In this case, “the match is scratched” together with true conditions C like sufficient oxygen is present, the match is dry etc. and the law “matches scratched in conditions C light” imply that “the match lights.” He immediately noticed that this works only if laws and C continue to hold were the match scratched. He called true conditions that would continue to hold “cotenable with the antecedent A.” Goodman assumed that laws are cotenable with A as long as A doesn’t violate the laws, but that other truths might not be cotenable with A.[[3]](#footnote-4) For example, if someone would not scratch a match in certain circumstances unless it was wet then in this instance being dry is not cotenable with “the match is scratched”. Specifying conditions cotenable with A involves further counterfactuals and so Goodman concluded that he couldn’t find a noncircular account of the truth conditions of counterfactuals. At this point he threw up his hands! This is “the problem of cotenability.”

This how the situation remained until the 1960s when both Stalnaker’s and Lewis proposed possible world semantics for counterfactuals.[[4]](#footnote-5) They found what looked like a solution to the problem of cotenability in terms of a similarity relation on possible worlds. According to these accounts the counterfactual “if A had been true at t then B would have been true at t” (A(t)>B(t’)) is true at world @ iff there is a pw W at which A(t) and B(t’) are true and there is no pw W\* at which A(t) and not B(t’) are true that is as similar to @ as W.[[5]](#footnote-6) They also claimed that “if A had been true at t then B might have been true at t’” (A(t)\*🡪B(t’)) if ff at least one world most similar to @ at which A(t) is true B(t’) is true.

Both accounts posited a reflexive, anti-symmetric and transitive similarity relation on the set of possible worlds. They differed in that on Stalnaker’s account this relation doesn’t allow for ties while on Lewis’ account there may be ties and there may not be a world at which A(t) is true that is most similar to the actual world but only increasingly more similar worlds without limit. (These differences are unimportant for the ensuing discussion.) These semantics support logics of counterfactuals that for the most part agree with intuitions.[[6]](#footnote-7) More importantly, they also suggested a way of solving the problem of cotenability in terms of similarity. The idea is that true statements that are also true in the most similar A(t) worlds are cotenable with A(t). The problem of cotenability is replaced by the problem of specifying the right similarity relation.

While these accounts support logics for counterfactuals their truth conditions depend on the nature of the similarity relation. Kit Fine pointed out in his review of Lewis’ book *Counterfactuals* that the right similarity relation is not overall similarity.[[7]](#footnote-8) Fine observed that overall similarity makes

(N) “If Nixon had pressed the button at there would have been a nuclear war”

false, since a world in which Nixon presses the button there is no nuclear war is more similar to the actual world than a world in which there is a war. What is the “right” similarity relation? Lewis’ intends his account to apply to worlds whose dynamical laws are deterministic and cover al physical events.[[8]](#footnote-9) If the dynamical laws are deterministic then similar worlds with a false antecedent compatible with the laws must be ones in which the laws are violated, or the microscopic histories differ or both. Lewis proposed a bit of both.

According to Lewis’s account three factors are relevant to determining the similarity relation relevant to evaluating counterfactuals. 1) conformity to the laws of the actual world 2) perfect match with the fundamental facts of the actual world 3) similarity with facts of the actual world.1 is most important, 2 less important and 3 is of little of maybe no importance. Since the laws are deterministic so it is not possible to completely satisfy 1 and 2 Lewis proposes balancing maximizing the length of times of perfect match in fundamental fact while minimizing the size of the region and the extent to which the fundamental laws are violated. The second is much more important than the first so minimizing the size of the violation of the actual laws is worth exchanging for large failure of match in particular fact. Lewis calls a violation of an actual world law at the counterfactual world “a miracle” even though it isn’t literally a miracle since it doesn’t violate a law of that world. Lewis’ way of balancing these two factors involves large and small “miracles”. His proposal is that in evaluating how similar w is to the actual world it is most important to avoid large miracles, second most important to maximize duration of perfect match, third to avoid small miracles, and fourth to maximize similarity but this last he says is of little importance.

Lewis’ account at first seems to agree with intuitions for a wide class of natural language counterfactuals. For example, it apparently evaluates Kit Fine’s example (N) as true assuming that the nuclear launch system was in working order and the laws are deterministic. It is plausible that the most similar worlds to the actual world at the time of the antecedent are ones that matches the actual world exactly until a moment when a small violation of laws of the actual world led to some neurons in Nixon’s brain to fire and to his pressing the button. Since the system was in working order conformity to the actual laws in these alternative worlds lead to a nuclear war. The counterfactual world conforms to the actual laws except for the small interval when Nixon’s neurons fire and matches the actual facts until just prior to that time but differs enormously with respect to particular facts afterwards. More important than getting ordinary language counterfactuals right, it seems to yield the counterfactuals that can be used in accounts of causation, influence, knowledge and so on.[[9]](#footnote-10)

Lewis distinguished the class of counterfactuals his account applies to from those he called “back trackers.” Backtrackers say how things would have had to have been in the past if A were true at a later time. For example, “if I were in Paris today, I would have had to leave NY yesterday.” Unlike backtrackers the counterfactuals connected with causation (C caused E) say that if an event C had not occurred at time t then E would not have occurred after t. These exhibit a temporal asymmetry from past to future. Lewis said his account doesn’t apply to back trackers, but we will see later that the probabilistic account can apply to some of them as well as to counterfactuals relevant to causation etc.[[10]](#footnote-11)

Lewis also did not intend his account to apply to counterfactuals like “If Caeser were a general in the Korean war, he would use an atom bomb (catapults)” or to counter legals or counterfactuals about mathematics. My probabilistic account doesn’t apply to them either and I know of no account that does.

Lewis claimed that the temporal asymmetry of causal counterfactuals is due to the fact that on his account small counterfactual differences at time t typically make for big differences after t but not before t. Accounting for this temporal asymmetry of counterfactuals would be enormously significant because the fundamental laws of physics are temporally symmetric.[[11]](#footnote-12) I mean by this the familiar fact that for every temporal sequence of macroscopic states compatible with the dynamical laws there is another sequence of macroscopic states compatible with the dynamical laws proceeding in the reverse temporal direction in which particles have the same trajectories that as far a fundamental dynamics is concerned is equally likely.[[12]](#footnote-13) If, as is widely believed, causation is not fundamental then it is a puzzle what accounts for its temporal asymmetry.

Lewis though that his account captures the temporal asymmetry of ordinary counterfactuals since typically a small miracle will get the actual to diverge at a time t to a world at which the counterfactual antecedent obtains while once the antecedent obtains a large miracle is required to get that counterfactual world to converge to the actual world. He observes that the reason for this is that in our world events typically have just a few pre determinants but leave many traces of their occurrences. For this reason, it takes a little miracle to yield the counterfactual event but a large miracle to wipe out all its traces in order to obtain convergence to the actual world. If his account was successful, it would provide the basis for the temporal asymmetry of causation for the other “arrows of time.” That would be a great achievement, one might even say “a miracle.”

Lewis’s possible world account seems to make whether a counterfactual is true or false depend on what is going on in non-actual possible worlds. But this is a bit misleading. On the account the truth/falsity of counterfactuals supervenes on the actual world. This is clear since according to Lewis laws supervene on the actual world and the size of regions and extent in which those laws are violated and the size of regions in which actual facts are violated depend only on the actual world. This means that the proposition expressed by a counterfactual is made true/false solely by the actual world. Lewis is well aware of this. He says:

It's the character of our world that makes some A-worlds be closer to it than others. So, after all, it's the character of our world that makes the counterfactual true - in which case why bring the other worlds into the story at all? To which I reply that [it] is indeed the character of our world that makes the counterfactual true. **But it is only by bringing the other worlds into the story that we can say in any concise way what character it takes to make what counterfactuals true.** The other worlds provide a frame of reference whereby we can characterize our world. By placing our world within this frame, we can say just as much about its character as is relevant to the truth of a counterfactual: our world is such as to make an (A-and-C)-world closer to it than any (A-and-not-C)-world is.

Possible worlds are used as a device for characterizing that proposition and the logic of counterfactuals. Lewis has his reasons for holding that possible worlds are as concrete as the actual world and including them in his ontology, but his account of counterfactuals is not a good reason.

Unfortunately, despite its initial attractions, Lewis’ account of counterfactuals fails badly. There are several problems. Reviewing them will lead to a better account that replaces similarity with probability.

The first problem is that unlike Goodman’s account, on Lewis’s account laws are not perfectly cotenable with antecedents consistent with the laws if determinism is true. This is because if A is false in the actual world the most similar worlds where A is true are ones which contain “miracles.” Lewis thinks this doesn’t create much of a problem since the miracle always occurs prior to the time of A and is small. We will see that this is a mistake. But even if Lewis is correct about this it is still a problem. As several authors have pointed out, his account validates counterfactuals like “if I had scratched my ear a few moments ago the fundamental laws of physics would not have held.” If demonstrating that the laws of quantum theory are false should garner a Nobel prize this would entail the truth of “If I had scratched my ear a few moments ago I should have won a Nobel prize.” That seems crazy. A better account is one that makes laws cotenable with any antecedent compatible with the laws. But as we saw if the laws are deterministic then the most similar counterfactual worlds to the actual world must differ from the actual world for all time. This also seems wrong since it seems to entail that if I had scratched my ear the prior history of the world would have been different all the way back to the big bang. Later, I will show how we can live with this.

A second problem has been extensively discussed by Alan Hajek. He points out that if we take seriously the physics of our world then most typical counterfactuals are literally false. What he has in mind is that according to quantum mechanics if I were to drop a rubber ball on the floor there is a very small chance that it doesn’t bounce but quantum tunnels to the apartment below.

This seems to imply “If I were to drop the ball it might not bounce” and that implies “If I were to drop the ball it would bounce” is false. The same holds for pretty much any counterfactual whose antecedent is a proposition concerning a physical event. There is a small chance that the event won’t occur.

Hajek also observes that even in classical mechanics, which is deterministic, as long as the antecedent event is macroscopic and so can be realized by many different micro events, there will be a small probability of the ball not bouncing. This is illustrated by the familiar example of there being a small probability of the air in the room spontaneously fluctuating so as to be concentrated into one of the corners of the room. In fact, given the macro state of a system at time t there is a small probability of the system fluctuating in accord with the fundamental dynamical laws so as to realize pretty much any macro property at all involving the same number of particles as long as energy is conserved. The reason is that there are an enormous number of possible microstates that might realize any macro state and some of these evolve in such unusual ways. These probabilities come from statistical mechanics.17

Unless Lewis adds to his account some condition that makes worlds in which unlikely events of quantum tunnelling or fluctuations less similar to the actual world (in which such events sometimes occur) than worlds in which the consequent is true it looks like it is committed to most counterfactuals being true.

The fact that Lewis’ theory can count counterfactuals with macroscopic antecedents and consequents as true might be thought to count in its favor since ordinary folk count them as true too. This would be so if the aim of the theory were to capture ordinary uses of counterfactuals. But if the aim is to capture the core counterfactual, then if is a defect. These are the counterfactuals employed in the sciences and show need to pay attention to science. Just as physics and other sciences have taught us that sometimes what has seemed obviously true (e.g., material objects are solid through and through) is false it has taught that most counterfactuals are false. Hajek says that we should gracefully accept this and devise an account of counterfactuals that respects it. Despite what physics says about the core counterfactual there have been many attempts to devise truth conditions for ordinary language counterfactuals on which the counterfactuals that ordinary language speakers assert can be true. This is the first project that I mentioned earlier. Hajek has argued that none of these attempts are successful.18 As far as I can see he is correct but my concern in this paper is not with the first project but providing an account of the counterfactuals that are needed in philosophy of science. This project must pay attention to physics. The lesson to take from Hajek’s discussion is that if we take physics seriously then only those true contingent counterfactuals whose antecedents concern events and must have consequents that are probabilistic. The counterfactuals involved in accounts of causation, influence, and so on are false unless their consequents are probabilistic.

Lewis’s account needs to be developed to include counterfactuals whose consequents are probability propositions. In fact, he does extend his account to cover indeterministic dynamics. This results in counterfactuals whose consequents are probability propositions. If the fundamental dynamical law specifies probabilities for how states of the universe evolve then it may be that “miracles” in alternative worlds are not needed to make a counterfactual antecedent true. It may be that at the time Oswald was about to fire at Kennedy there was a probability that the universe evolves so that he misses. But then after the antecedent is true there may be a small probability that this world evolves so as to match the actual world. That is there may be a small probability that while Oswald misses Kennedy dies as if struck by a bullet and the world goes on as our world did. Lewis tried to deal with this with his notion of “quasi-miracle” claiming that the occurrence of Kennedy dying as if struck by a bullet when in fact Oswald did not shoot him is quasi miraculous. He doesn’t provide conditions for an event being “quasi-miraculous.” It is not just having a very small probability. But I won’t pursue this further since we will soon look at an account that accommodates probabilities and has no need for quasi-miracles or even miracles whether dynamics is indeterministic or deterministic.

A third problem for Lewis’ account involves counterfactuals which while not exactly what Lewis calls “back trackers” involve some backtracking. Consider, for example, “If Biden had been in Moscow on May 13, 2022, it is likely he would have been arrested.” Lewis’ account seems to evaluate this as true since it would take only a small miracle to get Biden to board a plane a day before yesterday that takes him to Moscow. But we think this counterfactual is false since we think that if Biden were in Moscow yesterday it likely would have been because arrangements had been

made for his safe visit. To evaluate it we backtrack to an earlier time when a small miracle leads by law to arranging his visit so that Biden gets an invitation and safe passage. The backtracker “If Biden had been in Moscow yesterday, he likely would have been invited” seems true. One might  
respond to this and similar examples by saying that the earlier miracle is smaller than the later one but that doesn’t seem to be the case. Lewis doesn’t tell us how to measure the size of miracles but size of regions in which the violations occur seem about the same.

A fourth problem related to the preceding one is what Lewis’ account says about counterfactuals like

(J) If my jacket had been stolen sometime yesterday before midnight it would have been stolen just a few moments before midnight.[[13]](#footnote-14)

Lewis’ account looks like it evaluates this as true since it requires perfect match until the last possible moment at which a small “miracle” results in the jacket being stolen. But (J) doesn’t seem to be true since it is easy enough to imagine circumstances in which my jacket might have been stolen at any time between yesterday and midnight. Some of the times at which my jacket might have been stolen may be more likely than others. This problem is separate from the problem that consequents of true contingent counterfactuals have probabilistic consequents. It is due to the way the extent of perfect match plays a role in determining similarity.

The fourth problem for Lewis’s account is that there are worlds that differ in their pasts from the actual world but converge to it by a small miracle so that their futures match perfectly. Bennett already pointed out that there are worlds that like this and thought this calls Lewis’ idea of  
grounding temporal asymmetries on an asymmetry of miracles into question.[[14]](#footnote-15) Adam Elga showed Bennett’s suspicions are correct by showing that there are worlds in which the counterfactual antecedent is true and match the actual world perfectly from a time shortly after the time of the antecedent that are most similar to the actual world. That is, Elga showed how given the physics of our world for any counterfactual A(t)>B(t’) with both false  
antecedent and consequent if t’ is significantly after t there are A(t) worlds that converge to the actual world at which B(t’) is false if t > t. This makes A(t)>B(t’) false. This demonstrates that Lewis’ account fails to capture the temporal asymmetry of counterfactuals that he hoped would explain  
time’s arrows and it is devastating to his account.

Elga’s argument briefly is this. Consider the counterfactual A(t)>B(t’) where A(t) and B(t’) are both false and t’ is after t. On Lewis’ account if this counterfactual is true there is world that matches the actual world exactly until a short time prior to the time t when a small miracle results in A(t) and then leads by law to B(t’). All worlds like that must also be ones at which B(t’) is true. Take a state of the actual world far after t and reverse the velocities of all the particles in that world. It will match the future of the actual world until a time t’’ between t and t’ and by a small miracle diverges to make A(t) true and then continue by law. Call this world W. In W B(t’) is not true since it is not true in the actual world. Now reverse the time order of events in this world. This world, call it W\*, starts off very differently form the actual world leads to the truth of A(t) and then converges to the actual  
world a short time later. W\* is an A(t) world that is most similar to the actual world but at which B(t’) is false since the convergence occurs before t. This is a very odd world since in it there are no traces of A(t) after the convergence of it to the actual world. In this world entropy start off high and decreases until the time of convergence and then increases. The world will contain an apparent “record” that report that A(t) did not occur after t and also apparent “records” of events that occurred in the actual world even though they don’t occur there. A weird world indeed! It follows that on Lewis’ account A(t)>B(t’) is false. Of course, Hajek already showed this. But the situation is much more dire since Elga has shown that size of “miracles” won’t by itself capture the temporal asymmetry of counterfactuals.[[15]](#footnote-16)

These problems establish that Lewis’ similarity semantics fails. Before introducing my probabilistic alternative, I will briefly discuss another account  
of counterfactuals called “the simple theory” also based on similarity that avoids at least one of these problems. The simple theory is due to Jonathan Bennett and has recently been advocated by Kadri Vihvelin and Terrence Tomkow.[[16]](#footnote-17) The idea is that to evaluate A(t)>B(t’) take the actual world and  
find a world W that is as similar as possible to the actual world except that A(t) is true at W(t). This requires removing from the actual world the fundamental facts that are incompatible with A(t) and replacing them with facts that make A(t) true in a way that doesn’t violate the fundamental laws. The truth condition for the Simple Theory is:

A(t)>B(t’) is true iff B(t’) is true at all the A(t) worlds that most resemble the actual world at t.

The laws of the most similar world at which A(t) is true are identical to the laws of the actual world. On the simple theory laws are cotenable with the antecedents consistent with them. Because of this the simple theory doesn’t involve “miracles” and so doesn’t run into the problem of validating counterfactuals like “If I had scratched my ear, I would have won the Nobel prize.

To deal with Hajek’s problem advocates of the simple theory needs to extend to counterfactuals with probabilistic consequents. I am not sure how to do this so leave it to its advocates. The “Jacket problem” is still a problem for the simple theory but not for the same reason as on Lewis’ account since it doesn’t involve maximizing perfect match in history. As we will see shortly the problem is far worse due to a feature of the underlying physics that the simple theory ignores.

The problem that Elga pointed out doesn’t affect the simple theory since it requires that closest A-worlds resemble the actual world only at the time of the antecedent. If the laws are deterministic then the entire histories of these worlds differ from the history of the actual world. So, on the simple theory

“If A had been true then then the past would have been true all the way back to big bang” is true.”

Some of the A(t) worlds most similar to the actual world at t may have pasts very different from the actual world. But the simple theory is not intended to apply to backtrackers, so its advocates don’t consider this a problem. As far as anything said so far, the simple theory doesn’t seem to endorse any temporal asymmetry of counterfactuals. So, it doesn’t validate “if Nixon had  
pressed the button the past prior to that would have been pretty much the same while the future would have been radically different. This may not worry proponents of the simple theory greatly since they don’t think it is the job of an account of counterfactuals to ground time’s arrows.

At first, the simple theory looks pretty good. But a second look reveals that because of the physics of our world it also fails. The problem is caused by the fact that there are worlds which are macroscopically identical to the actual world but at which there is a statistical mechanical fluctuation that results in A being true but the world outside of the region of A is macroscopically just like the actual world. Some of these worlds resemble the actual world at t just as much as more usual worlds that contain fluctuations only rarely. Consider, for example,

“If a statue resembling Donald Trump had been in Times Square at noon on Jan 1, 2022, there would be no macroscopic traces of how it got there.”

The simple theory counts this as true since there are lawful histories that are macroscopically identical to the actual history and to the region outside of Times Square and so are very similar to the actual state but in which a statue resembling Donald Trump fluctuated into existence a moment before noon Jan 1, 2022.[[17]](#footnote-18) But surely this counterfactual is false  
since if a statue resembling Donald Trump had been in Times Square it would have gotten there in some way that would have been recorded by the surveillance cameras around Times Square which would still be present in the alternative worlds.

It is not just examples involving bizarre fluctuations that create problems for the simple theory. The counterfactual  
  
“If Biden had been in Times Square at noon on March 15, 2022, and attacked by a rabid Trump supporter he would have been defended by the secret service.”

is evaluated as false by the simple theory. The state of the world as similar as possible to the state at noon March 15. 2022 but which has Biden in Times Square being attacked by a rabid Trump supporter but there is no secret service since he decided to slip out of his hotel room on his own is more like the actual state at that time than one in which there is secret service in Time’s square. There is no need for a fluctuation to produce this state. While the simple theory avoids some of the objections to Lewis’s account it faces problems of its own and it doesn’t even try to provide an account of temporal asymmetries.

Before describing a better account of counterfactuals, it will be useful to briefly discuss why counterfactuals are so important to us. There are, I think, two main uses for counterfactuals in our thinking. One is in decision making. When considering alternative decisions, I consider what will happen given that I make a certain decision in certain circumstances. To find out I keep the circumstances fixed just prior to making the decision add the decision and the laws will say what will happen. I do it like this when I consider the decision “up to me” and so independent of the circumstances in which I make it. For example, what would happen if I now decided to jump off this ledge. The fact that I wouldn’t jump unless there were a net below me doesn’t mean that I should suppose that a net is below me. The reason that there is not a net is the reason I won’t jump.

Another use of counterfactuals is to consider what would happen if something occurred in the future or the past which will not or wasn’t necessarily brought about by anyone’s decision. In these cases, we are asking what would happen given a plausible account of how the antecedent came about. For example, what would have happened if a category 5 hurricane had struck Manhattan yesterday or what would have happened if Biden were visiting Moscow yesterday. To evaluate this counterfactual, we consider that it takes time for hurricanes to form a travel and that during that time preparation are likely to be made This counterfactual involves a certain amount of back tracking to a time when the antecedent is likely.

The problems we found with Lewis’ account and the simple theory are due to the physics of our world. But they can all be solved by an account of counterfactuals that pays close attention to that physics. To explain this account, I need first to explain the relevant physics.

As mentioned earlier, the fundamental dynamical laws that have been taken seriously by physicists are (or may be) temporally symmetric. This means, roughly, that if a sequence of fundamental events is compatible with the dynamical laws, then there is a temporally reversed sequence of fundamental events that is also compatible with the laws. In classical mechanics the time reverse sequence of states reverses the velocities of the particles. This sequence of states is also compatible with the laws. As far as the fundamental laws are concerned one kind of sequence is no more common or likely than its associated temporally reversed sequence.[[18]](#footnote-19) But, of course, we have never seen the reverse sequence. Our world is full of processes that seem to be temporally directed.

It is not just that there are processes that are temporally directed but that their temporal direction seems to be a matter of law. Chief among temporally asymmetric laws is the “second law of thermodynamics.” The second law (in one of its formulations) says that the entropy of an energetically isolated system never decreases and if the system is not in equilibrium increases.[[19]](#footnote-20) There are various (more or less extensionally equivalent) ways of characterizing entropy. Entropy measures the amount of energy in a system that is not available for work. In statistical mechanics the entropy of a system is, roughly, the number (or measure of the volume) of the microstates that realize the macro state of the system[[20]](#footnote-21). A system S’s macro state at a time is specified by portioning the system into small regions and specifying the compositions, volumes, temperatures, mass densities, average radiation intensity and frequencies, and perhaps other macro variables of each region.

Macro states (or macro histories) compose a partition of the space of physically possible microstates (histories) that are large enough to be epistemically accessible and that behave more or less lawfully. Thermodynamic states are like this. It must be granted that the notion of macro state is vague since the macro state of a system depends on the size of the volumes in the partition. But as long as the volumes are sufficiently  
small (but big enough so that each volume has a finite measure and so is realized by infinitely may possible microstates) the exact size is not significant for the purposes of our ensuing discussion.[[21]](#footnote-22)

The second law of thermodynamics is exemplified by the melting of an ice-cube in warm water, the diffusion of a gas, the clumping of matter by gravitational attraction, and so on. All these processes are governed by the fundamental dynamical laws and are also entropy increasing. But the  
temporally reversed associated processes which are entropy decreasing are also compatible with the fundamental dynamical laws. So, the question is what explains the second law if its violation is compatible with the fundamental dynamical laws? The issue is central to our concerns since all sorts of apparently temporally asymmetric processes relate to increasing entropy.

Part of the problem of accounting for the second law was solved by Boltzmann. He observed that in a system whose macro state is not one of maximum entropy there are “many more” microstates sitting on trajectories that realize non-decreasing entropy than realize decreasing entropy. “Many more” is not quite right since there are just as many- infinitely many- entropy decreasing as entropy increasing trajectories. The right way to put it (as Boltzmann did) is that the measure of states on entropy increasing trajectories on the usual Lebesgue measure is approximately 1 while the measure of those on entropy decreasing trajectories is approximately 0. Further, within  
small macroscopic volumes of phase space the measure of entropy increasing states is approximately 1 and the measure of states that evolve in ways that are typical is approximately 1. However, there are within the region states that are entropy decreasing and states that evolve in bizarre, atypical  
ways, for example, there may be compatible with the macro state of a drop of ink dissolving into water as microstate that evolves into the shape of an elephant, although the measure of bizarre states like this is 0.

Boltzmann understood this measure as or as determining a probability density over possible states. Since the dynamical laws are deterministic exactly how probability should be understood in this context is a problem. Boltzmann seems to have understood statistical mechanical probabilities  
as based on ignorance of the exact microstate but that they are objective since they are determined by the dynamics of such systems. This probability assumption entails that if S is an isolated system (consisting of many particles) not at equilibrium it is very likely (probability almost 1) that the entropy of S will increase. An instance of this is that it entails that it is very likely that an ice cube placed in an energetically isolated pail of warm water will melt. Boltzmann’s account reformulates the second laws as the probabilistic law that it is very likely that the entropy of an isolated system not at equilibrium will increase. The rate of increase depends on the macroscopic state of the system.

Boltzmann’s probability assumption shows how thermodynamics is connected to fundamental microphysics and how probabilistic version of the second law is reducible to fundamental physics. While Boltzmann’s probability assumption “solves” one problem it raises another. While the measure of the set of states compatible with the ice cube in warm water that are sitting on entropy non-decreasing trajectories towards the future is 1 the measure of states that ate sitting on entropy increasing trajectories toward the past is also approximately 1. Due to the temporal symmetry of the dynamical laws, there are just as “many” states whose lawful evolutions are entropy increasing toward the past as toward the future compatible with an ice cube sitting in warm water at t. This means that on Boltzmann’s probability assumption it is almost certain that the cube arose spontaneously out of water at uniform temperature. Of course, this is absurd.

One could deal with this particular instance of the problem by specifying that the macro state of the pail at the time it became isolated contained a larger ice-cube in warmer water, but this would still make incorrect or no predictions about times prior to that.[[22]](#footnote-23)One wants an account that applies not just to some subsystems but to all subsystems of the universe and to the entire universe from the time of the Big Bang. There is a solution to the problem that goes back to Boltzmann and was explicitly discussed by Eddington, Feynman and others.[[23]](#footnote-24) It has recently been elaborated and defended by David Albert.[[24]](#footnote-25) Albert proposes adding to the dynamical laws a law characterizing the initial (i.e., at or right after the “big bang”) macro state of the universe as one which has very low entropy and which is highly symmetrical (one region is very much like any other).[[25]](#footnote-26) He calls this statement “The Past Hypothesis” PH. It is the sort of description that cosmologists have in mind when they describe the universe immediately after “the big bang” as enormously dense plasma of elementary particles and fields occupying a very tiny region at a terrifically high temperature. Cosmologists have argued that the entropy of this state is very small when gravitational entropy is taken into account.[[26]](#footnote-27)

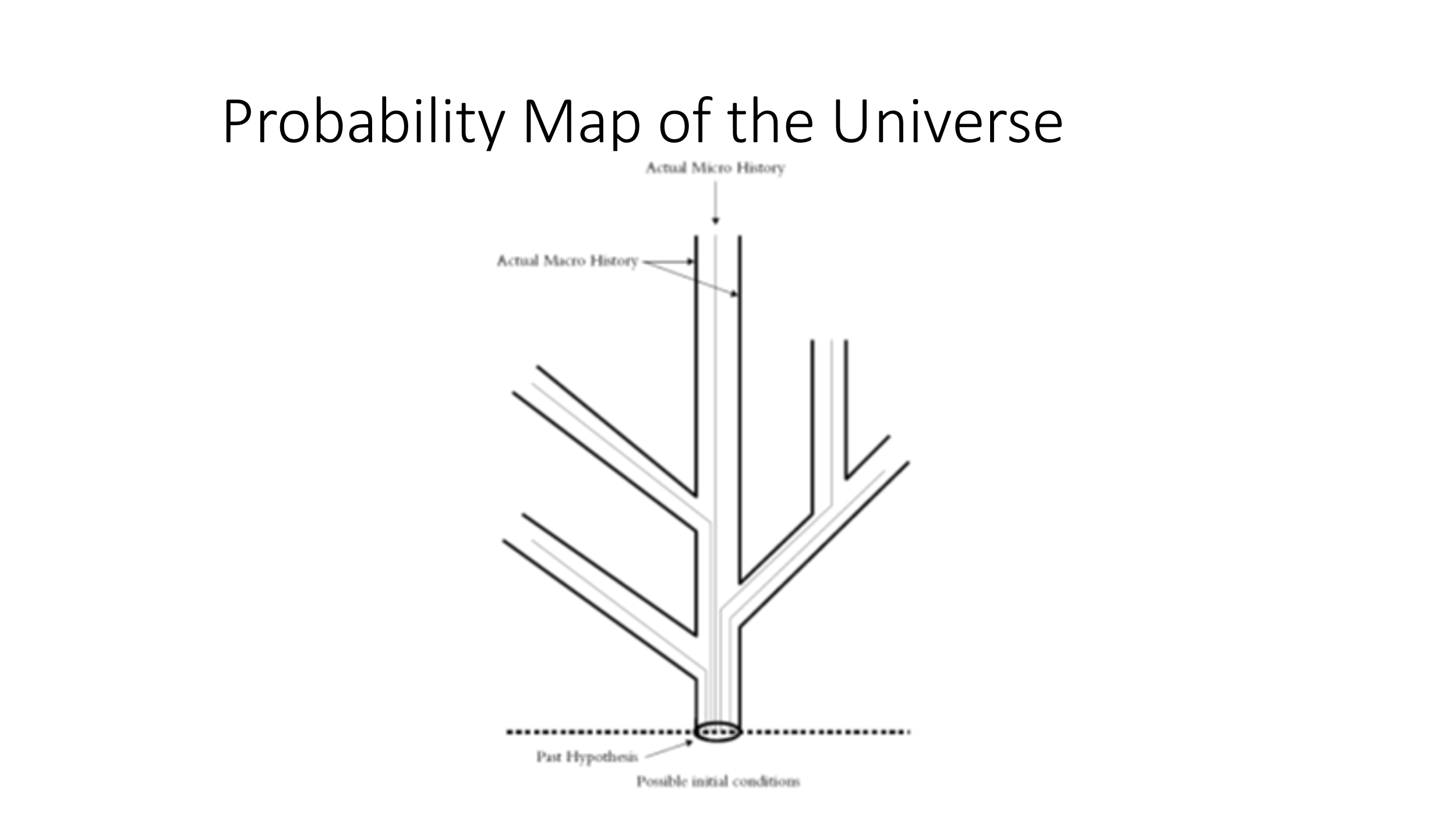
David Albert put the following three principles into a package that he and I call “the Mentaculus.”[[27]](#footnote-28)

1. The Dynamical laws (e.g., Hamilton’s equations for classical mechanics), Schrodinger’s equation for quantum mechanics
2. PH) a law characterizing the initial macro state of the universe as a very low-entropy condition satisfying certain further symmetry conditions. These further conditions characterize the initial macro state of the universe.

1. Prob) law that specifies a uniform probability distribution over the physically possible initial conditions at or very shortly after the origin of the universe.

Following Boltzmann, Feynman et al. Albert argues that the Mentaculus entails that at every time it is enormously likely that the entropy of the universe and everyone of its isolated subsystems increases in the temporal direction away from the PH. The important feature of the Mentaculus for our discussion of counterfactuals is that it entails conditional probabilities P(B/A) for all propositions B, A for which P(A)=/=0 and A and B are physical propositions i.e., they are equivalent of sets of fundamental physical histories. For example, it assigns conditional probabilities P(Knicks win the NBA championship in 2027/current macro state of the universe).[[28]](#footnote-29)

According to the Mentaculus even if the fundamental microphysical dynamical laws are deterministic, the macroscopic history of the universe evolves in deterministically. This is depicted in the following diagram



The thin lines represent micro histories, and the red line represents the actual micro history. The circle represents the special low entropy MACRO state postulated by PH. Of course, the circle occupies only a tiny portion of the space of nomologically possible initial conditions. The cylinders represent MACRO histories, and the blue cylinder represents the actual macro history. All the micro histories that initiate in the circle satisfy the low entropy initial condition and almost all evolve deterministically towards a state of maximum entropy. Although the evolutions of micro histories are governed by deterministic laws, MACRO histories appear to evolve in deterministically. That is, the MACRO state at t doesn’t determine a unique evolution but the probability distribution specifies the probabilities of the histories that realize that MACRO state. The branching of the cylinders represents the indeterministic evolution of macro histories. It is possible for macro histories to converge but this involves extremely rare microhistories. When macro histories converge records of many “past” events are erased.

The Mentaculus may seem to presuppose a direction of time since one of its components is called “The Past Hypothesis.” But this is misleading. The PH characterizes the macro state of the universe as having very low entropy at a particular time. But it doesn’t say that this time is in the future or in the past. Rather the PH earns it name by the role it plays in account for the temporal asymmetry of counterfactuals that explain the temporal asymmetries of influence, knowledge and causation.

The Mentaculus provides the probabilities needed by my theory of counterfactuals. The basic idea of the new account is that a true statement is cotenable with the antecedent A of A>P(B)=x iff C is probabilistically independent of A under certain circumstances M, i.e., P(C/A&M)=P(C/M).[[29]](#footnote-30) ). Probabilistic independence replaces similarity in the account of cotenability. This requires an account of where the probabilities come from and an account of the metaphysics of probabilities. The Mentaculus answers the first and I will return to the second at the conclusion of this paper.

The new account proposes these truth conditions for A>P(B)=x

A(t)>P(B)=x is true iff there is a time t\* between the time PH holds and t when the macro state is M(t\*) such that P(B/A(t)&M(t\*))= x and t\* is the time closest to t between PH and t for which P(A(t)/M(t\*)) is sufficiently large.

. The Mentaculus entails that t\* is almost always prior to t. At t\* the macroscopic state M(t\*) branches into alternative histories some which make A(t) true and on some of these B may also be true. [[30]](#footnote-31)

How large P(A(t)/M(t\*)) needs to be depends on context. If A(t) is a “free” decision, then it can be very small since we think of the decision as “up to” the agent and not the prior physical state even though the agent is a physical system.[[31]](#footnote-32) If the counterfactual is a causal counterfactual- e.g. if he hand not thrown the rock then the window would not have broken- the probability can also be very small so that given the macroscopic state a moment before the throwing there is a small probability that he doesn’t throw the rock. But in other cases context may mandate that the P(A(t).M(t\*)) is larger even as great as .5. The larger P(A(t)/M(t\*)) the more back tracking there will be. As mentioned earlier a very unlikely statistical mechanical fluctuation can result in particles forming the shape of a statue of Trump in Time’s square but for it to be reasonably likely the branch point for there to be a statue of Trump will have to sufficiently earlier so as to make an account of how it got there likely.

The context dependence of P(A(t)/M(t\*)) allows for a certain amount of vagueness in much the same way that vagueness of similarity does in Lewis’ account. In evaluating a counterfactuals, we may be more or less interested in how the antecedent could have come to be true and depending on how likely we think the account should be that will require more or less backtracking. This results in differences in the truths that are cotenable with the counterfactual’s antecedent.

The Mentaculus account makes no mention of “similarity” or “miracles.” Instead of attempting to solve the cotenability problem in terms of similarity or resemblance among possible worlds or the states of worlds at a time the Mentaculus account deals with it in terms of conditional probabilities.

For example, “If Oswald had not shot Kennedy on Nov 23, 1230 pm, then it is highly likely Kennedy would have run for a second term” is counted true since at t\* a few minutes prior to t there is a significant probability that Oswald wouldn’t shoot and at that time it is highly likely that Kennedy would run for a second term given that Oswald doesn’t shoot him and the rest of the macro state at that time. The conditions other than Oswald shooting that would have made it likely that Kennedy would run are plausibly independent of Oswald’s shooting.

The probabilistic account solves all the problems that beset Lewis’s and the simple accounts. Here is an enumeration of its solutions and a few other advantages of the account.

1. On the Mentaculus account the consequents of counterfactuals are all probabilistic and so it doesn’t run into the problems that led Hajek to argue that typical counterfactuals are false. If you were to drop the ball it very likely would bounce P(bounce/drop&M)= almost 1. One problem solved.

2. The Mentaculus account doesn’t involve violations of law (“miracles”), so it doesn’t run into the problems Lewis account does which does involve miracles. There is no time at which had I scratched my ear at a later time it would be likely that I should win a Nobel prize. Second problem solved.

3.The Mentaculus account doesn’t require perfect match until t\* so it does involve a certain amount of back tracking. The Mentaculus account requires provides a unified account. For any false A(t) the micro history of the world must be different for (A(t) to be true. So, the backtracking is all the way back to the big bang. But this is, contrary to Lewis’ fears, innocuous. But from time t\* to the time of the PH the macro history of these different micro histories is very likely to perfectly match the actual macro history. This backtracking is innocuous since it doesn’t provide any way of influencing the past in ways we care about or can notice.[[32]](#footnote-33) Of course, because of the indeterministic structure of the macro histories macro histories after t\* may differ greatly from the actual macro history and from each other as in the examples of the Kennedy and Nixon counterfactuals. The backtracking interpretation results from requiring that the antecedent is likely given the nearest prior macro state.

4.If my jacket had been stolen yesterday then there are various probabilities for the times in which it is likely to have been stolen.

5. Obviously the problem that Elga raised for Lewis’ account doesn’t arise for the probabilistic account. Due to the Mentaculus probability structure local counterfactual differences at a time typically can result in alternative probabilities for big differences in the future but not in the past.

6. Since all the Mentaculus counterfactuals have probabilistic consequents here is no need to distinguish “would” counterfactuals and “might” counterfactuals. Would counterfactuals approximate Mentaculus counterfactuals with high probability consequents and “might” counterfactuals approximate those with low probability consequents.

7. The account applies equally well if the fundamental dynamics is indeterministic as long as it supports statistical mechanics. This is very different from Lewis’ account which when he attempted to accommodate it to indeterministic dynamics led him to introduce “quasi-miracles.” The Mentaculus account has no need for miracles quasi or otherwise. Aside from reasons discussed earlier this doesn’t work. One possible indeterministic dynamics for our world is the GRW version of quantum mechanics. Albert has shown that the GRW theory entails the statistical mechanical structure of the Mentaculus. The only differences in our depiction of the Mentaculus is that there is no probability distribution over initial states compatible with the PH and micro histories branch with probabilities at the branch points. Every macro branching corresponds to a micro branching but there are micro branchings that do not lead to macro branchings.

8. There is a kind of counterfactual that creates problems for Lewis’ account and has received much discussion called “Morgenbesser Conditionals”.

With a slight amendment the Mentaculus account easily deals with them. An example of a Morgenbesser conditional is this:  
Chump tosses an indeterministic coin and, whilst it is in mid-air, calls heads. The coin lands tails, and Chump loses. His betting was causally independent of the coin’s fall. It seems right to say: ‘If I had bet tails, I would have won.’[[33]](#footnote-34)

The indeterminism involved in the coin flip can be either statistical mechanical with deterministic. Lewis’ account evaluates this conditional as false on either case. So does the Mentaculus account. But if we amend the Mentaculus account to

if t\*\* is the time of utterance A(t)>P(B)=x is true at t\*\* iff there is a time t\* between the time PH holds and t when the macro state is M(t\*) such that P(B/A(t)&M(t\*)&M’(t\*\*))= x and t\* is the time closest to t between PH and t for which P(A(t)/M(t\*)) is sufficiently large and M’(‘t\*\*) is the macro state at t\*\* except for those facts that are probabilistically dependent on A(t) at t\*.

In other words, to calculate the probability of B(t’) were A(t) to have been the case we conditionalize not only on A(t) and M(t\*) but also on that part of M(t\*\*) that is independent of A(t). In this example, since the coin’s landing heads is included among the records at t\*\* the probability that Chump would have won is very close to 1. Other Morgenbesser conditionals can be handled similarly.

There is one important issue left to discuss. What does probability mean in the Mentaculus? As mentioned earlier Boltzmann seemed to understand the probabilities he posited in his account of statistical mechanics as epistemic specifying the degrees of belief one should have about a system’s microstate given that one knows its macro state. But since these probabilities occur in laws, they must be objective and not merely degrees of belief. There are a number of ways of understanding statistical mechanical probabilities objectively. The best I believe is provided by David Lewis’ Humean account of laws and chances. On that account laws are certain generalizations entailed by the best scientific systematization of the totality of fundamental property instantiations throughout space-time. Lewis calls this totality “the Humean Mosaic.”

Here is how Lewis adds probability to the BSA. A function term, "P(E,t)”, denoting a function from propositions and times to real numbers between 0 and 1 is added to the language for candidates for best system to enhance informativeness while minimizing added complexity. Truth conditions for sentences containing the expression are then provided by specifying criteria for evaluating systems containing these sentences. The idea is that P(E,t) is true iff the best system of the world entails it. Here is an illustration of how this works. Consider a long sequence of the outcomes of coin tosses, e.g., <HHTHTTTHHTTHTHHTTT… >. A very informative description of the sequence specifying each outcome would be very complicated. A simple description, for example, “approximately half Hs and half Ts” would be much less informative. But the description “outcomes of independent trials with probabilities P(H)=.5, P(T)=.5” is fairly simple and much more informative. If it is the best system for the sequence, then in the sequence P(H/T)=.5.

Understood this way, chances are not fundamental features of events. Rather, a function expression specifying chances is introduced into theories that are candidates for optimal scientific systems to provide information about the mosaic. The proposition “the chance of E at t is x” says that the actual mosaic is such that its scientifically optimal systematization(s) entail that the chance of E at t is x. In this way, chances are given by BSA laws that specify chances.

Lewis applies this idea only to chances that are specified by a certain kind of indeterministic dynamical law. These are laws of the form “if h is the actual history of the universe or an isolated subsystem up until and including time t, the chance of s at t is x.” In all the examples of fundamental theories I know with indeterministic dynamical laws, the dynamical chances at t are determined by the state just at t. The history prior to t is irrelevant. Sequences that satisfy such laws are called “Markovian.”[[34]](#footnote-35) The history prior to t is irrelevant so I will consider laws of the forms Pt (s(t’)/s(t))= x where s(t) is the state at t and s(t’) the state at a subsequent time. The law may also take the form of a differential equation that describes the evolution of chances. Lewis’s proposal is to permit axioms of candidates for best system Σ to include dynamical laws of this form. For such theories, the possible histories of the universe exhibit a branching structure in which at each branch point the laws determine the chances of futures of that branch.[[35]](#footnote-36) The values of chances at a world are specified by the dynamical laws of the best system of the world. For example, the chance now of it raining tomorrow is the sum of the chances of each of the branches emanating from the macro state of the world now on which it rains tomorrow.

A great advantage of the BSA account of probability is that it is possible to modify it to apply to objective probabilities, whether dynamical laws are deterministic or indeterministic. This is accomplished by adding to candidate best systems an axiom specifying a probability distribution or density over initial conditions compatible with the dynamics. This axiom qualifies as a component of a best system in virtue of its adding informativeness and fit at little cost in simplicity. This is how to understand the probabilities that occur in statistical mechanics, the Mentaculus, and the Mentaculus account of counterfactuals.

Ordinary users of counterfactuals know nothing of the Mentaculus, or the truth conditions specified by my account, or very much about probability, so how do they manage to get correct most of the time? That is a good question. I am not attempting to answer it but rather my account explains what makes counterfactuals relevant to account of causation, influence, decision etc. true/false. Plausibly, ordinary speakers learn enough about the macro structure of the world and its laws to so that their uses are close to the Mentaculus account. But it is beyond the scope of this paper to explain how that works.

References

Albert, David (1992) *Quantum Mechanics and Experience*  Harvard University Press

Albert, David. (2000) *Time and Chance* Cambridge Mass: Harvard University Press

Bennett, Jonathan (1984) “Counterfactuals and Temporal Direction”,The Philosophical Review 93, (1984), pp. 57–91]

Bennett, Jonathan (2003) A Philosophical Guide to Conditionals

Oxford University Press.

Elga, Adam. (2000) “Statistical Mechanics and the Asymmetry of Counterfactual Dependence.” Philosophy of Science suppl. Vol 68: 313-24

Fine, Kit (1975) “Critical Notice of Lewis’ *Counterfactuals* [*Mind*](https://philpapers.org/asearch.pl?pub=682) 84 (335):451 - 458 (1975)

Fernandes, Alison (forthcoming) “The Branchpoint Proposal and the Role of Counterfactuals”

Fernandes, Alison. 2023. Time, flies, and why we can’t control the past. In Barry Loewer, Brad Weslake, and Eric Winsberg, eds., *The Probability Map of the Universe: Essays on David Albert’s Time and Chance*, Cambridge, MA: Harvard University Press. pp. 312–34.

Goodman, Nelson 1955) *Fact, Fiction, and Forecast*

Hajek, Alan (2014) “Most Counterfactuals are False”

Hajek, Alan (2014) “Most Counterfactuals are False”

Hájek, Alan. 2014. Probabilities of counterfactuals and counterfactual probabilities. *Journal of Applied Logic* 12(3): 235–251.

Hajek, Alan (2018) “Most Counterfactuals are Still False”

———. 2021. Counterfactual scepticism and antecedent-contextualism. *Synthese* 199: 637–659.

Ismael, Jenann. 2012. Decision and the Open Future. In A. Bardon (ed.) *The Future of the Philosophy of Time*. London: Routledge.

———. 2017. An Empiricist's Guide to Objective Modality. In Matthew Slater & Zanja Yudell (eds.) *Metaphysics and the Philosophy of Science: New Essays*. New York: Oxford University Press. 109−125.

Khoo, Justin (2016). “Backtracking Counterfactuals Revisited.” Mind 126, 503 (October 2016): 841–910

Leitgeb, Hannes. 2012. A probabilistic semantics for counterfactuals. *The Review of Symbolic Logic* 5(1), 26–121.

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

———. 1979. Counterfactual Dependence and Time’s Arrow. *Nous* 13: 455−76.

———. 1994. Humean Supervenience Debugged. *Mind* 103: 473–390.

Loewer, B. M. (1979). Cotenability and counterfactual logics. *Journal of Philosophical Logic*, 99-115.

Loewer, Barry. 2001. Determinism and Chance. *Studies in History and Philosophy of Modern Physics* 32: 609–29.

### Loewer, Barry. (2004) “David Lewis’s Account of Objective Chance*” Philosophy of Science.*

———. 2007. Counterfactuals and the Second Law. In Causation, Physics, and the Constitution of Reality. ed. Huw Price and Richard Corry. Oxford: Oxford University Press. 293−326.

———. 2012. Two accounts of laws and time. *Philosophical Studies*. 60: 115–137.

Loewer, B. (2020). The Mentaculus Vision. In *Statistical mechanics and scientific explanation: Determinism, indeterminism and laws of nature* (pp. 3-29).

Loewer, Barry (forthcoming) “The Consequence Argument Meets the Mentaculus”

Penrose, Roger (1999) *The Emperor’s New Mind*

Slote, Michael

Stalnaker, Robert C., 1968, “A Theory of Conditionals”, in *Studies in Logical Theory*, Nicholas Rescher (ed.), Oxford: Basil Blackwell, 98–112.

Tomkow , Terry (2013) The Simple Theory of Counterfactuals” https://tomkow.typepad.com/tomkowcom/2013/07/the-simple-theory-of-counterfactuals.html

### Tomkow and Vihvelin (2017 [The Temporal Asymmetry of Counterfactuals](https://philarchive.org/archive/TOMTTA-5) 2017 - philarchive.org

1. There is an excellent discussion of these two projects and how they can sometimes pull in different directions in  
   Hajek “Most Counterfactuals are Still False” [↑](#footnote-ref-2)
2. Goodman [↑](#footnote-ref-3)
3. An event doesn’t violate a law if it is consistent with the content of the law, but it need not be consistent with the law being a law. In that case it undermines the law. Humean accounts of laws permit such undermining/ [↑](#footnote-ref-4)
4. Robert Stalnaker (1968) David Lewis (1973) [↑](#footnote-ref-5)
5. I am assuming for most of the rest of my discussion that the antecedent and consequent are associated with times t and t’ where t can be before or after t’. [↑](#footnote-ref-6)
6. Loewer (1979) [↑](#footnote-ref-7)
7. Fine (1975) [↑](#footnote-ref-8)
8. Classical mechanics is like this as are some of the most reasonable versions of quantum mechanics. [↑](#footnote-ref-9)
9. Lewis developed an account of causation based on his account of counterfactuals. The central idea is that C(t) causes E(t’) if had C(t) not occurred E(t’) would not have occurred. While there are problems with this account, and it is unclear whether it can be developed into an adequate account it does establish a close connection between causation and counterfactuals. [↑](#footnote-ref-10)
10. There is an extensive discussion of backtrackers in Khoo (2016). The approach taken there is quite different from the probabilistic account developed in the current paper. [↑](#footnote-ref-11)
11. In the case of classical mechanics, where the state of an isolated system is given by the relative positions and momenta of its component particles the temporally reversed sequence is obtained simply by reversing the velocities of all the particles. [↑](#footnote-ref-12)
12. An exception to this is the GRW formulation of quantum mechanics . See Albert (1992) [↑](#footnote-ref-13)
13. This example is taken from Bennett (2003) [↑](#footnote-ref-14)
14. Bennett (2003 ) [↑](#footnote-ref-15)
15. Lewis believed that the size of miracles does explain the temporal asymmetry of counterfactuals because once a small miracle occurs the event described by the antecedent leaves many traces that would take a big miracle to remove. He is right that this is true as a matter of fact in our world but there is nothing about the dynamical laws that guarantees this. As ell see something needs to be added to entail that events typically have few pre determinants but many post determinants. [↑](#footnote-ref-16)
16. Bennett (1974), Vihvelin ( 2016) and Tomkow and Vihvelin (2013). Bennet first proposed the simple theory but in (2003) he renounced the simple theory for reasons much like those advanced in this paper and continues to maintain this position (private conversation). Vihvelin and Tomkow have revived it and are defending it. [↑](#footnote-ref-17)
17. By saying “an elephant fluctuates into existence I don’t mean merely that a mirage of an elephant fluctuates into existence as Terry Tomkow suggested but that particles fluctuate so as to take the shape of an elephant and move like an elephant. [↑](#footnote-ref-18)
18. Or, more vividly, run any motion picture backwards and it depicts a sequence of macroscopic events that are compatible with the dynamical laws.  
      
     [↑](#footnote-ref-19)
19. See Albert (2000) for statements of various characterizations of entropy and versions of the second law of  
    thermodynamics.  
     [↑](#footnote-ref-20)
20. This characterization of Entropy is Boltzmann’s. The relevant measure is the normalized Lebesgue measure of the set of microstates. [↑](#footnote-ref-21)
21. It is not implausible that many of the macro properties of a system that interest us - e.g., kinds of macro-objects, motions of macro-objects, and so on – supervene on (or come close to supervening) on the macro state of the system. [↑](#footnote-ref-22)
22. More generally, while the probabilistic assumption provides good predictions for how a system will evolve macroscopically it makes incorrect retro dictions if employed by itself with respect to the system’s past; for example, it retrodicts that it is more likely that a photograph looking like Bill Clinton emerged as [↑](#footnote-ref-23)
23. Among those who have developed this kind of account are Carlo Rovelli, Roger Penrose, and Sean Carroll. See  
    Carroll (2010) for a discussion.  
     [↑](#footnote-ref-24)
24. David Albert Time and Chance 2000  
     [↑](#footnote-ref-25)
25. David Albert (2000). This idea is present in Boltzmann and is assumed by many others (Sklar, Feynman, Lebowitz) but the status of the PH and the consequences of the assumption for matters other than the second law were seldom discusses prior to Albert’s book.  
     [↑](#footnote-ref-26)
26. Penrose ( 1999) calculated the entropy of the early universe and its entropy now and that the former is much smaller than its current value and much much smaller that its value when the universe reaches and equilibrium  
    state (if there is one). [↑](#footnote-ref-27)
27. The name is taken from the Cohen brothers movie “A Serious Man” in which a minor somewhat deranged character says that he is writing “s probability map of the universe” and calls it “The Mentaculus.”

    [↑](#footnote-ref-28)
28. Assuming that there is a more or less determinate set of micro trajectories on which the Knicks win the championship in 2027. [↑](#footnote-ref-29)
29. The idea of analyzing counterfactuals in terms of probability and understanding cotenability in terms of probabilistic independence is a natural one. It has been proposed previously by Leitgeb, Hajek, Albert, Loewer and Fernandes. To develop it one needs an account of the relevant probabilities, and this is found only in Loewer and Fernandes which are based on an account of statistical mechanics due to Albert. [↑](#footnote-ref-30)
30. It is for this reason that Alison Fernandes calls her account of counterfactuals “The Branchpoint Proposal.”. This account is very similar to one presented here in that it dispenses with similarity in favor of probability and is based on a probabilistic structure like the Mentaculus. The two accounts were developed independently although both were greatly influenced by David Albert’s *Time and Chance.* [↑](#footnote-ref-31)
31. The Mentaculus account of counterfactuals can be used to reply to Peter van Inwagen’s well-known “Consequence Argument” against the compatibility of free will and determinism See Loewer (forthcoming) [↑](#footnote-ref-32)
32. This point is made in Albert’s *Time and Chance* (2000) and argued for in detail in Loewer (forthcoming) [↑](#footnote-ref-33)
33. The example is taken from Slote (1978) There is an excellent discussion of Morgenbesser conditionals in Phillips (2007)  
     [↑](#footnote-ref-34)
34. Lewis’s account allows for non-Markovian laws and there have been proposals for fundamental laws that are non-Markovian, but I will ignore them. [↑](#footnote-ref-35)
35. . An example of such a law is the dynamical law for the evolution of quantum states proposed in the GRW theory. See Albert (1992) for discussion. [↑](#footnote-ref-36)