11

Counterfactuals and the Second Law

BARRY LOEWER

I will be discussing a kind of conditional (or counterfactual) typically expressed in English by subjunctive conditionals. Here are some examples: 'if I were to strike this match there would be an explosion', 'if there had been an explosion it is likely that I would have heard it', 'if the explosion had not occurred then the window would not have broken.' This kind of counterfactual is intimately connected with laws, explanation, causation, choice, knowledge, memory, measurement, chance, the asymmetry of past and future, etc. a veritable *Who's Who* of philosophically and scientifically significant concepts. Philosophers may disagree about the order of explanation among these items and counterfactuals but everyone ought to agree that we would make significant progress understanding them all if we had an account of what makes this kind of counterfactual

This chapter arose out of hours and hours of discussions between David Albert and myself about statistical mechanics, counterfactuals, chance, and related issues. It was originally intended to be a co-authored paper. But as it emerged I ended up writing it and after consulting with David we decided that it is best if it is authored by myself alone. The ideas in the paper (as will be made clear) are built upon the account of statistical mechanics that David develops so compellingly in Time and Chance (Albert 2000). However, the reader should be aware that although there is broad agreement between David and myself concerning the foundations of statistical mechanics there are differences in emphasis. In the book David makes some intriguing suggestions about how counterfactuals are connected to statistical mechanics but doesn't develop the account and at this point he is unsure that he would develop them in the way I do here. It goes without saying, although I will say it anyway, that my thinking on this topic is enormously indebted to David's work. Other philosophers who have connected counterfactuals and statistical mechanics in ways similar to my approach are Adam Elga (2000) and Doug Kutach (2003). My discussion is indebted to their work to discussions with them. I am also grateful to Frank Arntzenius, Craig Callender, Tim Maudlin, Mathias Frisch, Brad Weslake, and Laurie Paul for comments on various earlier incarnations of this paper. I am especially grateful to Frisch's penetrating criticisms of earlier versions of this paper.

statement true/false; what facts in our world (and worlds like ours) serve as the truth makers of these counterfactuals. One answer to the analogous question for certain regions of discourse is that their basic statements express fundamental facts that do not supervene on anything more basic. But whatever attractions this kind of view may have for some subject matters (e.g. spatial relations, values, consciousness, and modality) it is off the table here. More basic facts about laws, causation, probabilities, the distribution of fundamental properties, dispositions, our interests and beliefs, possible worlds or something else determine the truth-values of counterfactuals.[1]

The counterfactuals of interest are typically temporally asymmetric. If a counterfactual's antecedent mentions a nomologically possible and relatively local event (or situation) then if that event had (not) occurred subsequent events would or would likely have been very different but the course of prior events (typically) would have remained pretty much the same. For example, if Hitler had been assassinated in 1943 the subsequent course of the war and world history would have been very different but history prior to the assassination would have been pretty much the same. Explanation, causation, chance, choice, and memory are also temporally asymmetric. It is tempting to think that all these temporal asymmetries have the same source. But what can that source be? The issue of where temporal asymmetries come from becomes especially puzzling once we realize that the fundamental dynamical laws of nature are—or for all we know might be—temporally symmetric. There is no 'arrow of time' to be found in them.[2] Some philosophers suggest that time itself has an intrinsic direction or orientation that underlies the apparent passage of time and the temporal asymmetries.[3] But even if there is something to this suggestion it is not at all clear how time's supposed intrinsic direction connects with counterfactuals (and the rest) in a way that can explain the

[1] Not Everyone agrees (Lange, forthcoming). A referee remarked that it would be interesting to argue for a claim from more basic principles. I agree that it would be but will leave that for another occasion.

[2] More accurately there is no 'arrow of time' in the known dynamical laws that plausibly can account for these asymmetries. Certain processes involving elementary particles (e.g. neutral kaon decay) seem to involve temporally asymmetric processes as a matter of law. But no one seriously thinks that these rare processes can be the source of the temporal asymmetries of counterfactuals and so on.

[3] Maudlin, 'On the Passing of Time' (2005) argues that time possesses an intrinsic direction and suggests that this direction is connected with the direction of nomological explanation and causation.

FN:4 temporal asymmetries.[4] Of course, one could simply take the direction of time as given and stipulate a way of evaluating counterfactuals (or stipulate that causes must temporally precede effects or that memories can only be of the past etc.) that necessarily supports temporal asymmetries. Rather than a *metaphysical* or a *linguistic* account the project pursued in this paper is to develop a *scientific account of* temporal asymmetries. My approach is to ground the temporal asymmetry of counterfactuals (the direction of causation and the other temporal asymmetries) in contingent facts and laws. Although the fundamental dynamical laws are (or may be) temporally symmetric many non-fundamental and special science laws are temporally asymmetric. Most prominently, the second law of thermodynamics says, roughly, that the entropy of an isolated system (the universe as a whole) never (or almost never) decreases over time. Lewis says that he thinks that there is a connection between his account of counterfactuals and the second law but that he '... does not know how to connect the several asymmetries ... and the famous asymmetry of entropy'. (1986). The project of spelling out that connection has only just begun and will be continued here.[5] FN:5

11.1 Scientific and Metaphysical Background

In this section I will go over some of the metaphysical and scientific background that is needed for my discussion of counterfactuals.

In the following I assume that *physicalism* is true. I won't try to define 'physicalism' but take it that it entails that all contingent truths, including true counterfactuals and special science laws and causal relations supervene on fundamental physical truths including the laws of physics.[6] Character- FN:6 izing 'fundamental *physical* property/relation/law' is a problem that has

[4] Loewer (2003) argues that an intrinsic temporal direction is neither necessary nor sufficient for grounding the temporal asymmetry of counterfactuals and the others.

[5] Horwich and Sklar both suggest that there is a connection between counterfactuals and the second law (and statistical mechanics) and more recently Albert (2000) has developed the connection in ways that I will build on. A somewhat different account, connected to Albert's, is taken by Kutach (2003).

[6] Lewis (1999) defines *physicalism* this way: 'Among worlds where no natural properties alien to our world are instantiated, no two differ without differing physically.' A property is alien to a world if it is not instantiated there or at any nomologically possible world. While physicalism has many adherents it is opposed by those who think that some fundamental properties are *mental* and others who think that there are *emergent* macro properties or emergent macro laws.

led some philosophers to say that 'there is no question of physicalism'. However, in my view the project of fundamental physics is to find (or make plausible proposals) concerning the fundamental ontology and laws and I think there is reason to think that this project has met with some success. Physicists currently have an idea (or rather a number of ideas) of what the complete and correct physical ontology may be like. I will be sticking my neck out insofar as I will assume that there is a fundamental physical ontology and fundamental laws and that they are close enough to current proposals so as to make no difference to the line of thought that I will pursue in this paper.

In non-relativistic theories the totality of values of fundamental quantities at a time for an isolated system is the state of the system at that time. For example, in classical mechanics the state of a system consists of the (relative) positions and momenta of all the particles at a time. Fundamental laws come in two varieties. Static laws place constraints (or assign probabilities) on (to) possible states at a time. Dynamical laws place constraints on possible histories. They specify how the state of an isolated system evolves or the objective chances of its possible evolutions over time (as long as the system remains isolated). In classical mechanics the dynamical laws (e.g. Hamilton's equations) specify deterministically how the state evolves. While the ontology of the world and its laws are not now known it is very plausible that the correct account will in ways that are relevant to this paper not be too different from the current candidates. In particular, the correct account will need to account for the approximate truth of classical mechanics, statistical mechanics, and thermodynamics.

Counterfactuals and causation are closely connected so a word about how causation comes into fundamental physics is in order. Russell (1913) observed that the fundamental laws—he was thinking of the differential equations of classical mechanics but the same holds for Quantum Mechanics—specify how the whole state of an isolated system evolves (or the chances of possible evolutions) but don't specify which parts of the state at one time are causally connected to which parts of the states at other times.[7] He concluded 'The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the

FN:7

[7] Russell's point has recently been emphasized by Field (2003), Elga (this volume), and Kutach (2003).

monarchy, only because it is erroneously supposed to do no harm' (Russell, 1913, p. 1). I don't think that Russell meant that we should (or could) cease to use causal locutions but rather that causation is not among the fundamental relations and taking it to be fundamental leads to philosophical confusions. Given Russell's observation one can still maintain that causal claims may be true (or approximately true) and useful but if so their truth or approximate truth supervenes on the fundamental physical laws and truths.[8] Some philosophers (e.g. Lewis) think that causation can be analyzed in terms of counterfactuals (and other fundamental facts) while others think that an analysis of counterfactuals presupposes causation. Alternatively, it may be that neither can be analyzed without reference to the other but that they can be analyzed together in terms of fundamental physical facts and laws. Or it may be that although while there is no analysis of causal claims they nevertheless supervene on fundamental physical laws and truths I don't provide an account of causation in the following but I do propose an account of conditionals that doesn't appeal to causation and is thus suitable to play a role in accounting for causation, perhaps along Lewisian lines.

As I mentioned earlier, most candidates for the fundamental dynamical laws that have been taken seriously by physicists are temporally symmetric. By 'temporally symmetric' I mean, 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. For classical mechanics if one sequence of particle positions is compatible with the laws so is the reverse sequence. As far as the fundamental laws are concerned one kind of sequence is no more common or likely than its associated temporally reversed sequence. Classical mechanics is like this and so are certain versions of quantum mechanics.[9] So, for example, just as there is a sequence of positions compatible with Newtonian laws in which a ball rolling on a flat surface eventually stops while heating up there is also

[8] Not everyone agrees. Tooley (1987) thinks that there could a duplicate of a world like ours (at least if our laws are indeterministic) with respect to laws and the distribution of fundamental properties but which differ in causal relations. While it might be that our concept of causation allows us to conceive of scenarios as possible reflection I think shows that this is not a concept of causation that has relevance for science or is instantiated.

[9] If the physical state in classical mechanics is the positions and momenta of each particle then of course the time reverse of a sequence of such states simply makes no sense. The appropriate time reverse sequence of states reverses the velocities of the particles. Since, plausibly, velocities supervene on positions the fundamental states are the positions are the particles and for any sequences of positions compatible with classical mechanical laws the reverse sequence is also compatible with those laws.

a sequence in which a ball initially at rest spontaneously begins to move while growing cooler.[10] But, of course, we have never seen this happen. Our world is full of processes that seem to be temporally directed. And, it is not just that there are processes that are temporally directed but that the 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.[11] There are various ways of characterizing entropy. Boltzmann's account is that the entropy of a system is the number (or measure) of the microstates that realize the macro state of the system.[12] A system S's macro state is specified by its composition, volume, temperature, mass density, average radiation intensity and frequency, and perhaps other macro variables. More generally, the macro state of a system at time t is specified by partitioning the spatial region occupied by the system into small spatial regions and specifying the values of macro quantities (average temperature etc.) in each of these regions at t. The macro states (or macro histories) partition the space of physically possible micro states (histories) that behave more or less lawfully. The usual thermodynamic states are like this. Obviously, the notion of *macro state* is vague and there are many precisifications that would serve the purposes of statistical mechanics. With one addition that will be introduced in Section 11.3 it suffices for my purposes to assume that the partition partitions the space of possible microstates into very small volumes of positive measure.

The second law is exemplified by the melting of an ice-cube in warm water, the diffusion of a gas, the spreading of waves, the clumping of matter by gravitational attraction, and so on. All of these processes are entropy increasing even though they are governed by dynamical laws that allow for entropy decreasing processes. So the question is what explains the second law if its violation is compatible with the dynamical laws? The issue is central to my concerns since many temporally asymmetric processes are connected with increasing entropy.

[10] Or, more vividly, run any motion picture backwards and it depicts a sequence of events that are compatible with the dynamical laws.

[11] See Albert (2000) for statements of various characterizations of entropy and versions of the second law of thermodynamics.

[12] This characterization of Entropy is Boltzmann's.

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 or increasing (toward the future) entropy than realize decreasing entropy. 'Many more' is not quite right since there are just as many—continuumly 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. Boltzmann understood this measure as or as determining a probability over possible states. Since the dynamical laws are (assumed to be) deterministic exactly how probability should be understood in this context is a problem. I will have a little bit to say about that later.[13] This probability assumption entails that if S is an isolated system (consisting of many particles) it is very likely (probability almost 1) that the entropy of S won't decrease in time. More specifically, it entails that it is very likely that an ice cube placed in an energetically isolated pail of warm water will melt. But while this 'solves' one problem it raises another. The measure of the set of states compatible with the ice cube that are sitting on entropy non-decreasing trajectories *towards the past* is also approximately 1. That is, there are just as 'many' states whose deterministic evolutions are entropy increasing toward the past as toward the future. This means that on Boltzmann's probability assumption it is almost certain that the cube arose spontaneously out of water of uniform temperature. Of course, this is absurd.[14] We could deal with this particular 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.[15] Another reaction is not to take these probabilities so seriously and employ them only as an instrument for

[13] It is usual to understand statistical mechanical probabilities as degrees of belief (measures of ignorance).

[14] 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 a fluctuation out of dispersed molecules than that it was the result of someone taking a snapshot of the ex-president. Taking seriously this undermines or claims to know the very theories that predict it.

[15] Something along these lines is suggested by Reichenbach (1956). His idea is that the statistical mechanical probability assumption is applied to systems that branch off and become energetically

FN:16 making predictions about the future.[16] But this begs the question of why this procedure works so well and its working is apparently lawful.

There is a realist and foundational solution to the problem that goes back to Boltzmann. The solution is to add to the dynamical laws a boundary condition specifying the macro state of the universe. This condition–which is called 'the Past Hypothesis' (PH)–characterizes the macro state at or soon after the Big Bang origin of the universe. Cosmology suggests that this macro state is one which has very low entropy and which is highly symmetrical (one region is very much like any other).[17] The proposal then is to add to the fundamental dynamical laws the following two claims.

FN:17

(PH) a statement specifying the macro state of the universe at one boundary (which is assumed to be on with very low-entropy condition satisfying certain further symmetry conditions).[18]

FN:18

(PROB) a uniform probability distribution over the physically possible initial conditions compatible with PH; i.e. the initial macro state of the universe.

I take PH and PROB to be laws and I will later provide justification for so calling them. I will write P(B/A) for the conditional probability function PROB(B/A&PH&L); where L are the fundamental dynamical laws (which are assumed to be deterministic). P(B/A) is the statistical mechanical probability distribution (the 'SM distribution') of B conditional on A. It is a consequence of these assumptions that as long as the state of the universe is not in equilibrium (maximum entropy) it is enormously likely that its entropy will increase in the direction away from the PH and that its entropy decreases in the direction of the PH.[19]

FN:19

Figure 11.1 is a depiction of possible evolutions of micro and macro states that should provide an idea of how all this goes. The thin lines represent micro histories, the actual micro history being the straight line

isolated from the rest of the universe. One problem with his approach is that it leaves unexplained where these probabilities come from.

[16] Something along these lines is suggested by Leeds (2003).

[17] 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.

[18] These further conditions characterize the initial macro state of the universe.

[19] There is a caveat. Albert (2000, ch. 5) shows that Maxwell's Demon situations are compatible with PROB and PH and so if world happens to evolve into a Maxwell's Demon situation at t the probability that entropy will *decrease* give the macro state at t is high.

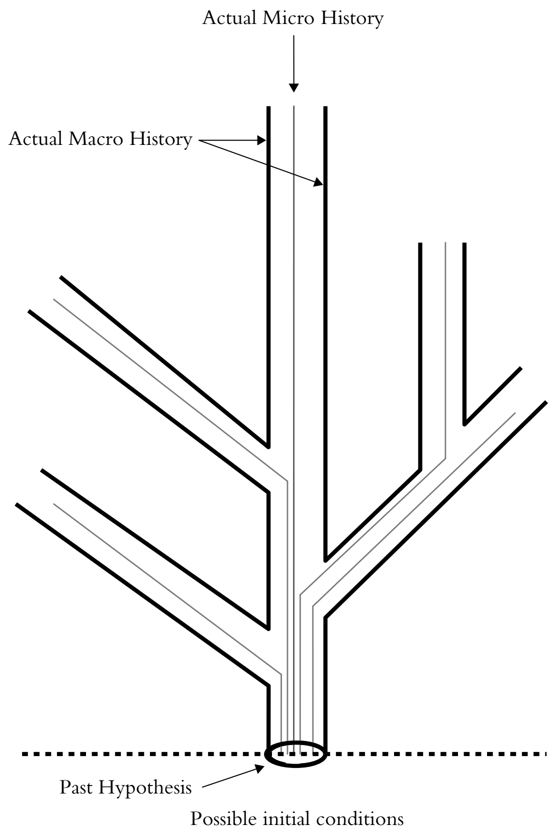

Figure 11.1. The world according to statistical mechanics

up the centre. The circle represents the special low entropy macro state postulated by PH. The circle occupies only a tiny portion of the space of nomologically possible initial conditions. The cylinders represent macro histories, the actual macro history being the cylinder that contains the actual micro history. All the micro histories that begin in the circle satisfy PH and almost all evolve towards a state of maximum entropy. However, there are a 'few' (i.e. sets of measure 0) maverick microstates whose entropy (i.e. the entropy of the macro state it realizes) does not increase, or increases for some time and then decreases, and so on. The interesting point is that although the evolutions of micro histories are governed by deterministic laws, typical macro histories appear to evolve indeterministically. That is, the macro state at *t* doesn't determine a unique evolution but the probability distribution

FN:20 specifies the probabilities of the histories that realize that macro state.[20] The branching of the cylinders represents the indeterministic evolution of macro histories. From a typical macro state in the middle of the actual macro history there will be branching in both temporal directions but there will be much more branching where the branches have substantial probability in the direction away from the time of the PH than back towards it. The overall structure is due to the fact that the macro state at t (in the middle) must end up in the direction of the boundary condition at which PH obtains (the direction we call 'the past') satisfying PH. There is no similar constraint at the other boundary condition (in the direction we call 'the future'). Of course, given the probability distribution and the dynamical laws it is very likely that the macro state at t will eventually evolve into a state whose entropy is maximum. But there are many many more *ways* in which the macro state at t can evolve to that condition than they can evolve from PH.

The SM probability distribution induces a probability distribution over thermodynamic propositions (propositions about the temperature, average mass distribution, etc. in small regions) and, more generally, over all FN:21 propositions that supervene on microphysical histories.[21] It is plausibly a consequence of the SM distribution that an energetically isolated (or approximately isolated) system not in equilibrium (an ice cub in a pail of warm water) will (as long as it remains isolated) evolve towards equilibrium (melted ice). Because the initial entropy of the universe was so low and is likely increasing, and because the macro states of isolated systems are typically not correlated with entropy decreasing (increasing) micro states, it is overwhelmingly likely that a system that becomes isolated and is not at FN:22 equilibrium will evolve towards equilibrium.[22]

The SM probability distribution embodies a way in which 'the future' (i.e. the temporal direction away from the time at which PH obtains) is 'open' at least insofar as *macro states* are being considered. Since all histories must satisfy the PH they are very constrained at one boundary condition

[20] The probability assignment over initial conditions together with the deterministic laws and the supervenience principles connecting macro to microstates determines a probability distribution over propositions characterizing macro states. So for example, it assigns a probability to the proposition that a particular ice cube in warm water will melt in the next hour.

[21] I am assuming that macro propositions are equivalent to sets of micro propositions and that the SM distribution induces a probability distribution over them.

[22] See Albert (2000).

but there is no similar constraint at other times. It is true that (almost) all histories eventually end up in an equilibrium state (there is a time at which almost all histories are in an equilibrium state) but this is not a constraint it is a consequence of the dynamics and the PH and it is not very constraining (almost all states are equilibrium states). Another feature of the SM distribution when applied to the macro state of the kind of world we find ourselves in is that the macro state of the world at any time is compatible with micro states that lead to rather different macro futures. For example, conditional on the present macro state of the world the SM probability distribution may assign substantial chances both to its raining and not raining tomorrow. On the other hand, there is typically much less branching towards the past. The reason is that the macro states that arise in our world typically contain many macroscopic signatures (i.e. macro states/events that record other macro states/events) of past events but fewer macroscopic signatures of future states/events. Newspapers are much more accurate in recording past weather than in predicting future weather. Of course these two features of the SM distribution—that histories are very constrained at one boundary condition but not at other times and that they branch much more to the future (direction away from the PH)—are related.

Albert shows how the assumption of the PH (and to consequent branching structure) allows for the production of localized macro records of past events while there are not comparable 'records' of future events. For example, a footprint on the beach at t2 is a record of someone having walked on the beach earlier. Albert points out that the inference from the footprint to the walker assumes something about the state of the sand (e.g. that it was soft and damp) at an appropriate earlier time. Without the assumption of the state of the beach at t1 (or something that plays that role) we are not justified in making the inference. More generally, Albert says that a localized macro state R at time t is a record of an event E at time t' if it can be inferred from the laws of physics and the macro state S at a time t'' (where t' is between t and t'') that E occurred (or probably occurred). He thinks of S as a kind of 'ready condition' that makes the production of the record possible. In the example, the sand's being damp is the ready condition. The PH plays the role of a ready state for our universe that allows the production of records. If PH were dropped then retrodictions from local current macro states (or from the entire) would be wildly inaccurate.

In view of the above the SM probability distribution together with the temporally symmetric dynamical laws determines a temporal asymmetry with respect to the probabilities of macro histories of the universe. As long as the macro state of the universe is not at equilibrium it is very likely to increases in entropy in one temporal direction and decrease in entropy in the other direction. To be clear, nobody claims (and it is not true) that the PH and the dynamical laws are themselves sufficient to account for the existence of recording systems let alone the particular records that have been formed in our world. What the PH does is to remove an obstacle to there being the accurate local macroscopic traces of conditions prior to t by serving as, in Albert's words, 'the mother of all ready conditions'.

Before continuing I want to set aside one possible misunderstanding. It might be thought that the SM distribution already assumes a future/past distinction since it is characterized in terms of 'the past hypothesis'. But it doesn't. It does assume that a certain probability distribution holds over all micro states and that this probability distribution has a temporal asymmetry built into it. But this does not assume that one of the boundary conditions is temporally prior to the other. That one of these but not the other is the temporally earliest condition is something that has to be earned. It is partly earned by the fact that entropy increases in one direction. But this is only the beginning of an explanation of temporal asymmetries that ground the distinction between past and future. To the extent that the other temporal arrows can be explained in terms of the probability distribution that grounds the asymmetry of entropy increase the PH will earn the title 'the *past* hypothesis'.

What is the metaphysical status of PROB and PH? There is a great deal of controversy concerning how to understand probabilities in statistical mechanics. The usual view is that they are 'ignorance' probabilities; i.e. the degrees of belief that one ought to have concerning the initial conditions of a system given that one knows that the system satisfies certain constraints (e.g. the system's temperature). There is little discussion of the status of PH but it is generally construed as a *merely contingent* (i.e. not law-like) statement. I think that both of these interpretations are mistaken. PROB and PH (together with the dynamics) are best considered to be laws and the SM-probabilities to be objective. The reason is that they underwrite many of the asymmetric generalizations of the special sciences especially those in thermodynamics and these generalizations are considered to be laws. PH and PROB can bestow lawfulness on them only if they themselves are laws and if PROB is a law then the probabilities it posits must be objective; not merely degrees of belief. It is interesting that Lewis' Humean accounts of laws is friendly to the proposal that PH and PROB are laws. I don't have space to go into detail here but the basic idea is this. According to Lewis a law is a contingent generalization entailed by the Best System of the world. The Best System is that true theory couched in a vocabulary that includes a conditional probability function, mathematical notions, and terms that represent fundamental properties (Lewis' calls them 'perfectly natural properties'). What makes it *BEST* is that it best combines simplicity, informativeness and fit. Adding PH and PROB to the dynamical laws results in a system that is only a little less simple but is vastly more informative than is the system consisting only of the dynamical laws. Lewis' official view about objective probabilities is that they are dynamical transition chances. But his account of laws and probabilities actually makes it easy to see how probabilities can also be assigned to initial conditions as PROB does.[23]

There are two further reasons that count in favor of counting PH and PROB as laws. One is that, as I will explain later, these hypotheses play a role in evaluating counterfactuals that is similar to the role that laws play. Second, if PH is taken as a non-lawful contingency then relative to the statistical mechanical probability distribution it is enormously improbable. This observation has led Price (2003) to think that this 'enormously improbable' initial condition requires some explanation.[24] It is easy to find this very puzzling since obviously no causal explanation is possible for an initial condition. The problem vanishes if PH is a law since it then constrains the probability distribution so that its probability is 1. PH and PROB are no more in need of explanation than is any other law.[25] My view is that

[23] On Lewis official account simplicity and informativeness are measured relative to the language whose predicates correspond to what he calls 'perfectly natural properties'. And on his official account of chance objective chances different from 1 and 0 are incompatible with determinism. Our proposal involves measuring informativeness relative to the thermodynamic language as well as the fundamental language. Once that is done it is a natural consequence that there are macro laws and objective deterministic chances. Loewer (1996) defends Lewis' account of laws, Loewer (2004) defends Lewis' account of objective chance, and Loewer (2001) argues for the claim that statistical mechanical probabilities can be understood in terms of Lewis' account of chance.'

[24] For a discussion of this issue see Price (2003) and Callender (2003).

[25] Of course if there is an explanation of PH and PROB from some dynamical or other laws that would all be to the good and completely compatible with Lewis' account of laws and our accounts of statistical mechanics and counterfactuals.

the fact that Lewis' account of laws naturally counts PH and PROB as laws and his account of probability provides an objectivist account statistical mechanical probabilities provides strong support for Lewis' accounts.[26]

What is the epistemological status of PROB and PH? As is the case for any fundamental law (or general scientific hypothesis) there is no question *of proving* PROB and PH. They are to be thought of as *conjectures*. The question then is whether so far they have yielded correct predictions and what further investigations can be carried out to 'test' them. The first point to note is that the Boltzmann probability assumption when used to make predictions about the future of thermodynamic systems has enormous support within statistical mechanics. But without PH the probability assumption is clearly false (and indeed inconsistent since it can only apply at one time). PH overcomes that problem in a way that plausibly preserves statistical mechanical predictions while avoiding the problem of higher entropy past. Further, the assumption of a low entropy and symmetrical macro state at the origin fits in well with what cosmology says about the early universe and so has independent support. I think these are strong reasons to take PH and PROB very seriously. Also, it appears that PROB provides correct objective probabilities in situations that are not obviously connected with thermodynamics. For example, it is plausible that when we conditionalize the outcome of a coin toss on the macro state (or on the macro state in the vicinity of the coin toss) at the time of the toss it is equally likely that a flipped ordinary coin will land heads as that it will land tails. The reason is that very small differences in the micro states (i.e. initial position of the coin and its initial momentum) compatible with the macro state associated with the coin toss lead to differences in the outcome and it is plausible that the SM distribution assigns equal probability to the 'heads' and 'tails' resulting states.[27] Similarly, it is plausible that PROB conditionalized on the relevant macro state provides correct probabilities for various chemical, biological, meteorological phenomena. The fact (as Albert argued in Albert (2000) and I am arguing in this paper) that PROB and PH can ground the existence of local records, the inferences we make about the future and the past, inferences from one part of the present macro state (one record) to the existence of another, and the account of

[26] For further discussion of these points see Loewer (2004), Callander (2003), and North (2004).

[27] Strevens (2003) contains an illuminating discussion of how macro probabilities emerge from micro probabilities and dynamics.

counterfactuals that I will later sketch all provide strong reasons to take this account with the kind of seriousness appropriate to any very general account of the world that has so far been scientifically successful and renders coherent relations between fundamental laws and ontology and aspects of, in Sellars' expression, the 'manifest image'.

For all the reasons above I take the system of the world based on PH, PROB, and the dynamical laws very seriously. But it must be admitted that a lot more needs to be done to make it persuasive that all objective probabilities—the probabilities involved in gambling devices, natural selection, and so on—ultimately derive from the SM distribution by conditionalizing on appropriate facts. However, it must be admitted that on further investigation it might turn out that the correct probability distribution is not PROB (but one close enough so that it also grounds thermodynamics) or that there is no single objective probability distribution or any at all. But, and this is emphasized, there is as of now, as far as I know, no evidence that disconfirms PROB and PH and a lot of intriguing support for them.

Here is a summary of the physics background: The problem that confronted Boltzmann is that the fundamental dynamical laws of our world are temporally symmetric but the laws of non-equilibrium thermodynamics, and in particular the second law, is temporally asymmetric. The problem is solved in a way that avoids the reversibility objection by adding PH and PROB as laws. These hypotheses go far beyond that in that they determine an objective probability distribution over all nomologically possible micro histories (and *a fortiori* over all macro histories and all macro propositions). Even though the underlying micro dynamics is deterministic macro-histories form a tree structure branching towards the future (away from the time at which PH holds). This suggests that the temporal asymmetry of counterfactuals and all the other temporal asymmetries connected with it are grounded in the temporal asymmetry of the statistical mechanical probability distribution.

There is another account of statistical mechanical probabilities that I want to mention briefly. It has been suggested (Albert 2000, chapter 8) that the GRW version of quantum mechanics grounds statistical mechanical probabilities. His idea is that the random GRW jumps are much more likely to move a system not at equilibrium onto an entropy increasing than an entropy decreasing trajectory since the measure of the former is so much greater than the latter. Thus statistical mechanical probabilities can be reduced quantum mechanical probabilities. If PH is added to GRW we obtain a branching tree structure that is like the tree depicted in Figure 11.1 (although, of course, the possible micro histories will also form a branching structure unlike the deterministic dynamics depicted in Figure 11.1).[28]

FN:28

11.2 Lewis' Truth Conditions for Counterfactuals

Let's get back to counterfactuals. My interest in counterfactuals is not the same as the linguist or philosopher of language who aims to construct an account of the semantics and pragmatics of the natural language sentences that express counterfactuals. That project faces many daunting difficulties. English counterfactuals are expressed in a number of grammatically different ways, there are many kinds of conditionals, counterfactuals are vague, they are plausibly context relative, they have Gricean implicatures and so forth. The semantics and pragmatics of ordinary counterfactuals is a messy matter. Lewis' approach, which I will follow, is to ignore most of these difficulties. The justification for this is the assumption that there is a conditional that works more or less like core cases of counterfactuals and that is centrally related to causation, laws, and the direction of time, choice, and so forth. It is this conditional that I am aiming to articulate. Whether this approach is justified will depend on whether there is a satisfactory account of a conditional that plays this role. I will use $A > B$ for the counterfactual conditional that I am after. On Lewis' classic account $A > B$ is true iff either there are no possible worlds at which A is true or there are A&B worlds that are more similar to the actual world than any world at which A & $-B$ is true. Soon after Lewis published his account Jonathan Bennett and Kit Fine objected that it yields results at odds with counterfactuals we readily accept.[29] Fine's example is this. 'If Nixon had pushed the button (say on 1 January 1973) there would have been a nuclear war' apparently is evaluated as false by the account since a world in which Nixon pushes the button but somehow there is a failure in the missile launching system

FN:29

[28] We could also add to this account a probability distribution over initial conditions but almost any initial distribution will be soon 'washed out' and the probabilities converge on the usual statistical mechanical ones.

[29] Fine (1975) and Bennett (1974).

is more similar to the actual world than a world in which there is a nuclear war. Lewis responded in his paper 'Time's Arrow' that Fine and Bennett were thinking of 'similarity' as something like 'overall similarity' but that it is similarity only in certain special respects that is relevant. The questions, of course, are what are these respects?

Before discussing Lewis' answer I want to describe a more recent although less ambitious proposal due to Jonathan Bennett.[30] Bennett suggests that, when considering worlds in which A happens at time *t*, the worlds that are most similar to the actual world @ are ones that are very similar to, or perhaps even identical to, @ from its first moments until some time shortly before *t* when (assuming A(*t*) is false) they diverge from @ and evolve in conformity to the laws so as to make A(*t*) true. Bennett calls the point of divergence 'a fork' and the segment of history leading up to A(*t*) a 'ramp'. He says that while the fork may involve a region in which the laws of @ are violated it would not *appear* at a macroscopic level to involve anything astonishing and in, particular, would not be taken by someone who had access only to macroscopic facts to involve a violation of the laws that hold at @. The fork should occur as late as possible so as to minimize gratuitous differences between the counterfactual world and @. Bennett's view can be pictured as follows:

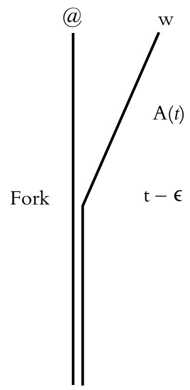

Figure 11.2

I think that Bennett's account does a pretty good job of characterizing a conditional that matches core uses of the counterfactuals that interest us.

[30] Bennett (2003).

For example, it seems to validate 'If Nixon had pushed the button there would have been a nuclear war' and, generally, it underwrites the temporal asymmetry of the kind of counterfactuals we are investigating. However, Bennett's procedure for evaluating counterfactuals *assumes* the distinction between past and future (since forks are to the future) and so it does not provide a scientific explanation of time's arrows.

Lewis' account does not build temporal asymmetry into his characterization of similarity and so holds the promise of explaining (rather than assuming) the temporal asymmetry of counterfactuals. His initial account assumes determinism. In evaluating which of two worlds is more similar to the actual world, four considerations in order of importance are relevant.

(i) The occurrence of big and widespread 'miracles' (violation of laws of the actual world.[31]
(ii) The size of the region in which the fundamental facts fail to match exactly.
(iii) The occurrence of small and very local 'miracles'.
(iv) The size and extent of dissimilarities with respect to fundamental (and other?) facts.

The last of these plays no role (Lewis expresses doubt in its relevance) in my discussion and so I will drop it. The three remaining factors determine a family of similarity relations depending on the relative weights attached to each.[32] Lewis suggests that the vagueness and context sensitivity of counterfactuals corresponds to indecision and shiftiness of the weights credited to these factors. Here is how he thinks the account applies to Fine's example. He considers four worlds at which Nixon pushes the button as candidates for most similar to the actual world @:

(1) w1: worlds which match the actual world exactly until a short time before the button pushing when there is a 'small miracle'—a small violation of @'s deterministic laws that leads to a decision to push the button and then by the laws of the actual world to the button being pushed and to a nuclear war.
(2) w2: worlds that conform perfectly to the actual laws but (since we are assuming determinism) differ from the actual world for all times and so over a large region.

[31] It may be that not only the size of the region but the extent of the violation is also relevant.
[32] Lewis says that the third of these is 'of little or no importance.'

(3) w3: worlds like w1 but at which a small miracle after Nixon pushes the button lead to a state that although isn't an exact match to the actual history (from a time shortly after the button pushing) is very similar to the actual world.

(4) w4: worlds like w1 but at which a miracle after Nixon pushes the button leads by the actual world's laws to exact convergence with the actual world.

Lewis claims that all the legitimate ways of weighing the three considerations for evaluating world similarity count w1 as more similar to @ than w2 since in w1 the region of match is so much greater than in w2 while the region of violation of the actual laws in w1 is small. He also claims that w3 is less similar to @ than w1 since the small difference between

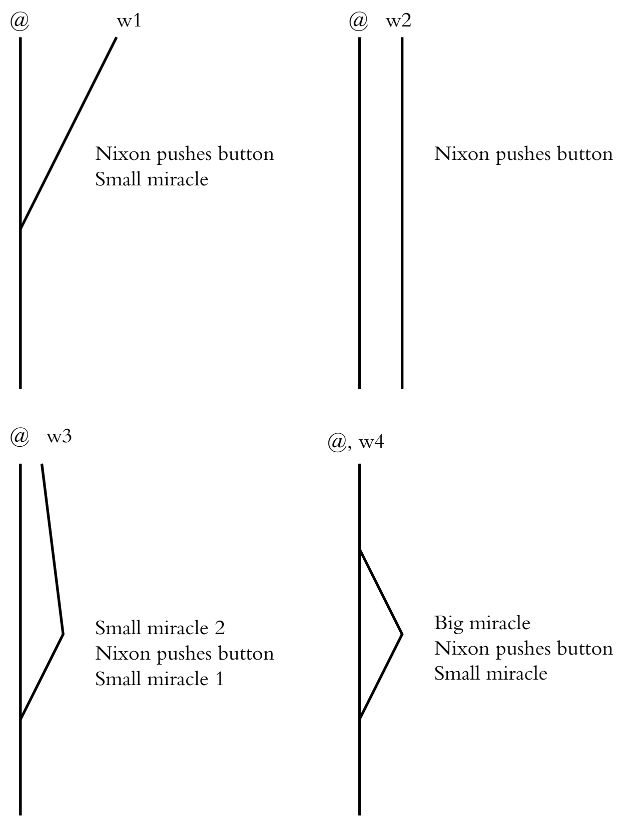

Figure 11.3

w3 and @ at times shortly after *t* inevitably amplify to great differences at later times. w4 would count as at least as similar to @ as w1 if the size of the miracle required to obtain exact convergence were small. But Lewis claims that certain features of @ require that the violation of the actual laws that occurs after *t* must be very big and widespread to bring about exact convergence. So w4 loses out to w1 in the contest for similarity. He claims that as a consequence Fine's conditional is evaluated by the account as true.[33]

Lewis thinks that the temporal asymmetry of counterfactuals is explained in terms of the asymmetry of 'miracles'; it takes only a small miracle for a world to fork from @ but a big miracle is required to achieve exact convergence to @. His reason for thinking that exact reconvergence requires a large violation of law is that he thinks that in @ there is an asymmetry of 'over determination'. He says that events at a later time over determine events at an earlier time but not *vice versa*. In the example, it is plausible that Nixon's pushing the button at t is compatible with the state of the world outside of Nixon's brain just prior to t. Once he has pushed the button there will be many dispersed traces in the form of radiation of various kinds that within a second occupy a sphere with a radius of 186,000 miles. This suggests that it takes only a 'small miracle' to result in Nixon's pushing the button but a large miracle is required to erase all the traces of that button pushing. Popper describes another vivid example. A rock is tossed into a quiet pool producing concentric waves. The rock's hitting the water is over determined by the many waves but the waves are determined only by the rock's hitting the water. Popper's idea is that an event e (or at least the kind of event that is referred to in an ordinary counterfactual) has many widely dispersed effects f1, f2, f3, ... that together

[33] I have heard it said that since according to Lewis the truth conditions of counterfactuals invoke possible worlds their truth makers involve other worldly events, laws, and so on. One hears the objection 'What does the fact that a match (even if it is a counterpart) lights in some other world have to do with whether *this* match in *our* world would light if struck?' But the objection is a mistake since it is a mistake to think that on Lewis' view the truth makers of counterfactuals are other worldly items. On his account the *truth makers* of counterfactuals are actual fundamental physical facts and laws. The truth maker of a (non-trivial) true counterfactual is a very complicated (and highly disjunctive) fact about the laws and fundamental physical facts. Possible worlds (and the similarity relation) are a device for characterizing those facts but are not part of the counterfactuals truth maker. Of course Lewis does believe that there are many causally disconnected universes (his possible worlds). But someone who thinks that possible worlds are properties or sets of propositions or fictions or even someone who takes modality as primitive can still accept his account of the *truth makers* of counterfactuals.

with the laws determine it–that is, each of the fs nomologically imply e but over determination of future by prior events is relatively rare. If this is what he means by 'determine' then (assuming determinism and laws like those of classical mechanics) he is wrong. For a localized event e(t) and time t' there will be a unique nomologically necessary and sufficient condition f(t') for e(t). Further, f(t') will typically be highly non localized; not the sort of condition we think of as an event at all. However, I think Lewis' intuition is correct when he says that events in our world typically leave many traces but have few predictors. Underlying this is the SM distribution and its role *in allowing* for the production of local macroscopic traces. So Lewis was onto something when he suggests that the asymmetry of over determination may have something to do with 'the famous asymmetry of entropy'.[34] To bring out the role of the SM distribution I will first discuss a problem for Lewis' account that has recently been highlighted by Adam Elga (2000).

FN:34

Elga argues that Lewis' account fails to ground the temporal asymmetry of counterfactuals and so delivers incorrect verdicts for the counterfactuals that interest us. On his account the temporal asymmetry of counterfactuals is supposed to be grounded in the temporal asymmetry of 'miracles'. But one may wonder where this temporal asymmetry of miracles can come from when the fundamental laws are temporally symmetric and considerations of perfect match are indifferent as to whether the regions matched are in the future or the past. Bennett seems to have worries along these lines. He mentions that there may be a world u to which another world v converges by a small violation of the laws of u. 'Easy convergence' is at least a possibility. But this doesn't show that Lewis' account is mistaken since a world to which easy convergence is possible may be very unlike @. But the existence of a world in which a typical local antecedent (e.g. Nixon pushes the button) that converges to @ via a small violation of law would be devastating for Lewis' account. That there are such worlds was pointed some years ago by David Albert and has recently been nailed down beautifully by Adam Elga.[35] Elga's argument goes like this. Let the actual world @ have a first and last moment and let @* be the time reverse of

FN:35

[34] Lewis (1986) remarks 'I regret that I do not know how to connect the several asymmetries I have discussed and the famous asymmetry of entropy' (p. 51)

[35] Albert made this point in a seminar at Princeton in 1996. Elga spells out this problem for Lewis' account in (Elga 2000).

FN:36 @.[36] The dynamical laws of @* are the same as those of @. @* evolves in conformity with the laws until a time t^* *prior* to the time t (the time at which Nixon is supposed to push the button). At t^* there is a small violation of the laws which leads to a fork on which Nixon pushes the button at t; i.e. there are particles that are so located so that they constitute a Nixon counterpart whose hand is on a button connected to nuclear missiles and so on. Call the resulting world w6. Now let w6* be the temporal reverse of w6. W6* has a past very different from @. It starts off in a state of high entropy. As it evolves towards time t entropy decreases as particles in it arrange themselves into the shape of Nixon pushing the button. Other particles and fields are so located so that all it takes is a small miracle a time a little after t to get this world to match exactly the actual world; no nuclear war, no traces of Nixon's button pushing, no traces at all of w6*'s past but rather 'records' of the history of @. It then evolves just as the actual world towards higher entropy. w6* is a strange world but it conforms to the actual dynamical laws everywhere except for a small region.

If the laws are restricted to the dynamical laws (as Lewis thinks) then Lewis' account of similarity counts w6 at least as similar to @ as w1. It matches @ in all of its future a short time after t and conforms to the laws

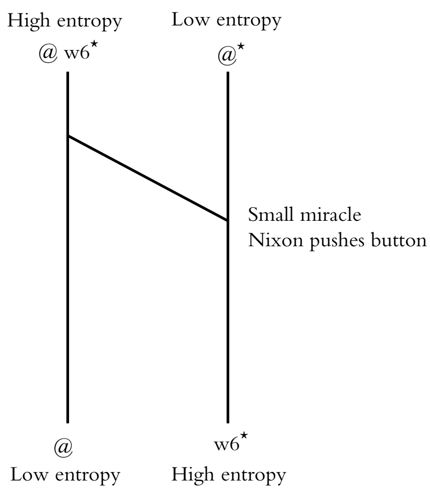

Figure 11.4

[36] In classical physics this means that the positions of particles at @*'s first moment matches @'s final particle positions and scalar field values but velocities and field vectors are reversed.

of @ except for a small region. This has disastrous consequences for Lewis' account. It means that his characterization of similarity fails, contrary to his belief, to underwrite the temporal asymmetry of counterfactuals and so fails to account for the counterfactuals connected with causation, decision, and so on. The problem is that @* is an anti-thermodynamic world. Where as in @ the entropy of isolated systems (almost) never decreases in @* entropy (almost) never increases. W6* is like @* in that initially it is in a very high entropy state (like @*) and evolves towards lower entropy until, by virtue of a small miracle, it matches the actual world and then evolves as it does towards higher entropy.

In light of the discussion of statistical mechanics this problem for Lewis' account should come as no surprise. The thing to do to repair Lewis' account is to add PROB and PH. The simplest way of doing that is to stipulate that conformity to PH and PROB are important in evaluating similarity to the actual world. The offending words @* and w6 will then count as less similar to @ than is w1. Adding them in this way provides another reason (in addition to their simplicity and informativeness) to PH and PRB as laws since they play the role of laws with respect to counterfactuals. This solution should be especially pleasing to those who favor Lewis' Humean account of laws since it counts PROB and PH as laws in virtue of their inclusion in the Best Theory of the world. It is not at all clear that alternative accounts can make sense of restrictions, especially probabilistic restrictions, on initial conditions as laws. I call Lewis' account that includes PH and PROB as laws 'the amended Lewis account'.

The amended Lewis account is not open to the Albert–Elga objection but there are still problems with it. It is plausible that there are worlds that satisfy the past hypothesis, match the actual world exactly until a short time prior to Nixon's pushing the button and then diverge by a 'small miracle' from the laws of the actual world and then reconverge by another small miracle to the actual world; i.e. worlds like w4. Lewis is correct in thinking that almost all worlds are not like that but general statistical mechanical considerations suggest that there are worlds in which a small violation of the actual laws lead to Nixon's button pushing as in w1 but in which unlike w1 no records are produced. If so only a little miracle is required for reconvergence. I can't prove that there are worlds like this (genuine proofs in statistical mechanics of complex systems are hard to come by) but the worry is that there are trajectories in phase space that satisfy the second law for a long time and then realize anti-thermodynamic behavior. Given PROB and PH the set of such worlds has a very tiny measure but the existence of any such worlds would be a big problem even for Lewis' account amended. If there is a single world that is like w4 then on Lewis' account the Nixon counterfactual will be evaluated as false. The remedy is to evaluate counterfactuals in a way that connects that more closely to the SM distribution.

FN:37

11.3 The SM-Account of Counterfactuals[37]

In this section I characterize a conditional—I call it the SM—that is temporally asymmetric, behaves in certain ways like Bennett's and Lewis' conditionals, and tracks the statistical mechanical probability distribution. The SM conditional that I will presently describe is one of many conditionals that respects the probabilistic structure induced by PBOB, PH, and the dynamical laws. It is offered here in a tentative way as part of an effort to see whether it captures that structure in ways that connect with the core counterfactuals that are themselves connected with causation, choice, and memory and so on.[38]

FN:38

Central to the account of SM conditionals are a special class of conditionals that involve decisions. These conditionals have the form 'If P were to decide to A then the probability of B would be x.' I make two assumptions about decisions. First that they are localized events (or states) in a person's brain that are smaller than macroscopic events but have positive probability and second that are correlated with motions of her body. For a given macro state of a person's brain the decisions are 'open' to her are those decisions that are compatible with the macro state of her brain. If determinism is true and our then the microstate of the world prior to my making the decision determines what decision will be made. But let's suppose (perhaps this is

[37] A different account of conditionals and their relationship to statistical mechanics is developed in Kutach (2003). I don't compare the accounts here but will do so in a subsequent paper.

[38] At this point Albert and I slightly part company. He (as have I) insists that any procedure for evaluating counterfactuals will need to respect and reflect PROB and PH. But he has been steadfastly agnostic in print about exactly what that procedure might look like. In talks and seminars, Albert has sometimes been willing to entertain counterfactuals with determinate consequents, but as we will see the account presented here is restricted to counterfactuals with probabilistic consequents. There are a few other places where my account sticks its neck out in ways that go beyond what Albert and Albert and I have written and said in talks.

merely a *fiction*) that even if determinism is true a person can directly control which decisions she makes. Further, I will suppose that given the macro state of the world (including the agent's brain) at the various decisions that are available to her are all equally likely.[39] Decisions are thus indeterministic relative to the macro state of the brain and environment prior to and at the moment of making the decision. This *indeterminacy* captures the idea that which decision one makes is 'open' prior to making the decision.

Now suppose I am choosing between alternative decisions D1(t') and D2(t'). If I knew the statistical mechanical probabilities P(B/Mt&D1) and P(B/Mt&D2) for various Bs I would be in a position to know the *objective* probabilities of D1 and D2 each leading to B. SM conditionals express information about these objective probabilities. Here is how.

Consider future subjunctive decision conditionals of the following form:

($) If at t I were to decide D1 then the probability of B would be x.

My proposal is that ($) is true if $P(B/M(t)\&D(t)) = x$.[40] For example, the conditional 'if I decide to bet on the coin's landing heads then the chance I would win is 0.5' is true at t iff the statistical mechanical probability of winning given the t macro state and my decision is 0.5.[41]

Decision conditionals are temporally asymmetric. They inherit the temporal asymmetry of the SM distribution. Alternative decisions that can be made at time t typically can make a big difference to the probabilities of events after t (i.e. events further away from the time at which PH holds) but make no difference to the probabilities of macro events prior to t. The reason for this is that the predominance of local macro signatures of the past (but not of the future). It also depends, of course, on our biological structure on which very small differences in the brain get *magnified* into differences in bodily movements and these, in some cases (e.g. the Nixon example) get magnified into vast differences in the world. So it is enormously plausible that decision conditionals are temporally asymmetric.

[39] The assumption that each possible decision is equally likely is certainly false but I don't think this simplification affects the following account.

[40] The reason it is we are interested in evaluating the conditional in terms of P(B/M&D) is that the *macro state* is a natural limit on the extent of accessible information and for typical conditionals this probability will be approximately equal to P(B/D&K) where K is a very small part of the *macro state*. In out example, for a typical coin flip everything but the nature of the flip is irrelevant to the probability.

[41] Gambling devices are designed so that generally the macro state external to the gambling device itself is irrelevant to the probabilities of outcomes. But, of course this isn't always correct.

Whether I conditionalize on D1 or D2 then the probabilities of macroscopic past events *insofar as they are recorded* in the present will be unaffected but the probabilities of the macroscopic future may very well be very different.

Here is a worry. Suppose that K is a statement that an event k occurred prior to time t and there is no macroscopic signature of K at t so P(K/M(t)) is very small and P(–K/M(t)) is high. For example, k might be the event of the destruction of Atlantis (supposing it to have occurred and left no macroscopic traces). Then P(K/M(t)&d) will be very small for all decisions (since these neither destroy nor create records) as well. This might lead one to think that the account says you could eradicate Atlantis by deciding to lift a finger. But of course this is not so. By lifting your finger you do not alter the probability of K at all. On the other hand you, or rather the president, by lifting his finger at t might well destroy the world at a time after t.

Here is another worry. It is a consequence of this account that the probability of the past micro state is correlated with present alternative decisions. But this does not mean that a person can affect the past in the sense of having control over past micro events. Control by decision requires that there be a probabilistic correlation between the event of deciding that p be so and p being so and one's knowing (or believing with reason) that the correlation obtains. But it is immensely implausible that there is any past micro state *m* that fulfills the first part of this condition let alone both parts. So while it is true on the account that if Nixon had pushed the button the probability that the past would have been different in some micro respects is 1. But since we have no idea what those respects are we have no control over past micro conditions.

Note that the temporal asymmetry of decision conditionals is derived from the asymmetry of the SM probability distribution since the asymmetry of local macro signatures is derived from the SM probability distribution. It is not 'put in by hand' as it is in Bennett's account. The asymmetry of these conditionals is part of the explanation of why our decisions can make a difference to the future but not the past and that partly explains our feeling that the future is 'open'.

Decision counterfactuals are a very special kind of conditional. What about conditionals whose antecedents are about a non-actual past decision and conditionals whose antecedents are not decisions? Consider, for example, 'If I had decided to flip this coin at noon yesterday the probability it would have landed heads is 0.5.' The obvious suggestion is that to evaluate it we look at the macro state M(noon). If the probability of the coin's landing heads conditional on M(noon) and my deciding to flip it is 0.5 then the conditional is true. It is plausible that the probability that Kennedy served out his term given the macro-state when Oswald decided to pull the trigger is pretty high. It remains high conditionalizing on the micro-event of Oswald's deciding at the last moment not to pull the trigger. If so the SM conditional 'if Oswald had decided not to pull the trigger it is likely Kennedy would have served out his term' is true.

So far I have discussed conditionals whose antecedents are decisions. This enabled us to keep the macro state fixed while altering the microstate so as to realize the alternative decision. Because the macro state can be preserved these alternatives don't 'back track' macroscopically.[42] But what about conditionals whose antecedents are macro events? Let's start with conditionals whose antecedents are bodily motions or actions; for example, 'if Nixon had pushed the button at t there would have been a nuclear war'. The natural way to extend the account to these conditionals is to find the latest time t^* prior to t at which it is open to Nixon to decide to push the button and which he immediately makes this decision and evaluate the conditional in terms of the probability of there being a nuclear war conditional on Nixon deciding at t^* to push the button and pushing it at t. This procedure will evaluate 'if Nixon had pushed the button at t there very likely would have been a nuclear war' as true given the usual assumptions about the circumstances of the button pushing.[43] This counterfactual antecedent does backtrack a bit. If Nixon had pushed the button at t the probability that he would have decided to do so at t^* is near 1. But this doesn't undermine the asymmetry of choice since it is only via decisions that we can affect bodily movements.

What about antecedents that are not actions? Consider for example, 'if a fire had started in the forest at noon it would have very likely destroyed

[42] This isn't quite correct. There may be past macroscopic events that leave no records in the present macro state and are very unlikely relative to the present macro state. For example, suppose that there was an Atlantis that left no traces. The probability of there having been Atlantis conditional on the current macro state would then be very tiny. So on our prescription if I had decided to lift my left hand (instead of my right) it is unlikely that there would have been an Atlantis. This is a strange consequence but not really unsettling since it doesn't entail that we can have control over past events.

[43] Of course, if the button was not connected or if Haig had arranged for Nixon's brain to be monitored so that should he decide to push the button it becomes disconnected we would obtain a very different result.

all the houses by the next day'. The natural extension of the account is to find a time t^* prior to t at which there are micro historie which diverge from the actual macro-history, and also to a fire starting in the forest at noon. If P(houses destroyed the next day/M(t^*)& fire started at noon) is high, the counterfactual is true. We can see why we might be interested in this probability and the corresponding counterfactual. If P(houses destroyed/M(t^*)& fire) is high, while P(houses destroyed/M(t^*)& no fire) is low, then whether or not the houses are destroyed depends on whether or not there is a fire. So whether or not there is a fire provides (or would have provided) a kind of 'handle' on whether or not the houses are destroyed. We are interested in fire histories that diverge from the actual macro history because the divergence point (what Bennett calls 'the fork') is the size of a decision and it is decisions over which we have direct control. As a 'first stab' at an account I propose:

(\$⋆) 'If A(t) had been true then the chance of B would have been x' is true iff t' is the latest time at which a divergence from the actual macro history similar in probability to a decision event can occur and P(B/A(t)&M(t')) = x.[44]

FN:44

The fact that my proposal for evaluating non-decision conditionals makes use of the distinction between past and future it doesn't undermine my claim that the asymmetry of counterfactuals is grounded in the SM probability distribution. The temporal asymmetry emerges from decision conditionals and depends on the SM distribution. So we have an explanation in terms of physics—a *scientific* explanation—of why we can affect the future but not the past. The proposal for evaluating non-decision conditionals is parasitic on decision conditionals.

There are problems with (\$⋆). In the example, there may be many ways in which fires can be started by 'small' events occurring a bit before noon some of which lead to the houses being destroyed and some not. The trouble is that the antecedent is under specified. Different additions to the antecedent—the fire starts in the north corner etc.—may lead to different conditional probabilities. One way to handle this would be to average over

[44] Requiring that the divergence event is similar in size (i.e. in probability) to a decision rules out enormously unlikely 'fluctuations' that can occur at later times.

the ways in which the fire could start using the probabilities of the small events that lead to fires at *t*'. Another difficulty is that there may be times prior to the latest time at which it is much more likely that a fire starts. Perhaps, a camper threw away a lighted cigarette a bit before *t*' that didn't start a fire but came close. In that case I am inclined to think that if a fire had started it would have likely been at this earlier time in the vicinity of the thrown cigarette. Perhaps some sort of averaging over an interval prior to t' would help here.

There is also a worry about backtracking. If Nixon had pushed the button at t then the probability that at some time t' prior to t a decision size event would have occurred that evolved the state by the laws so that Nixon pushed the button at t. There is similar backtracking on Lewis' account. I don't see this as especially worrying, since it doesn't at all undermine the claim that we cannot control the past. However, certain other backtrackers are more troubling; for example, 'if there were a nuclear war at t then the button would have been pushed at t'. The trouble is that this suggests backward causation Perhaps this can be handled by characterizing causal connection in terms of probabilistic correlations of the right sort and causal priority in terms of the temporal direction of control by decisions. But I leave the job of characterizing causal connection and causal priority for another occasion.

Even though I have not presented a complete account of counterfactuals it may be worth comparing it with Lewis' account.

(1) The SM account is not a similarity account like Lewis' but rather it is based on probabilities provided by the laws. However, in those cases to which both accounts apply the SM account plausibly gives very nearly the same result that Lewis thought his account gives and that Bennett's account does give. On the SM account it is plausible that 'If Nixon had pushed the button at t it is very likely that there would have subsequently been a nuclear war' comes out as true. Same for 'If Nixon had pushed the button at t it would have still been likely that ...' where ... is filled in by any macro past event.

(2) The SM account avoids the problem of 'converging' histories that Albert and Elga pointed out. Lewis can take a step towards avoiding this

problem as well by requiring that similar worlds are ones at which PH obtains. But it is only a step since Lewis' account still allows that there may be some worlds like w4 but in which it only takes a small miracle to yield convergence to the actual world. Such worlds contain very, very unlikely fluctuations but it is not clear that the dynamical laws rule them out. If there are such worlds then as I pointed out Lewis' account is sunk. The cause of the trouble is the fact that perfect match count as one of the criteria of similarity. The SM account doesn't work like that and so doesn't have this problem.

(3) The SM account goes some distance towards explaining the intuitive appeal of the idea that in evaluating counterfactuals we look at alternatives that are very much like the actual history until a short time prior to the time of the antecedent. But where Bennett simply declared this and Lewis tried and failed to account for it in terms of his similarity criteria the SM account *explains* how the laws of physics and the centrality of decisions underlie this way of evaluation.

(4) The SM account explains how the temporal asymmetry of counterfactuals is connected to the second laws of thermodynamics. Both are grounded in PROB and PH.

On the other hand, Lewis' similarity account is more general than the SM account. His account applies to counterfactuals that have both probabilistic and non-probabilistic consequents where the SM account applies only to the latter. It must be granted that ordinary language counterfactuals generally do not have probabilistic consequents. There is work yet to do to connect the SM account to ordinary language counterfactuals. Also, the SM account depends on the particular laws (and specifically PROB and PH) that obtain in our world. Lewis' account doesn't depend on the particular laws in the actual world. Perhaps the best way of looking at the relationship between the accounts is that while Lewis (and Bennett) were offering something close to an *analysis* of counterfactuals that is supposed to validate our intuitive judgments the SM account proposes a conditional that tracks the SM distribution in an especially interesting way. I say a bit more about this in the next section.

11.4 Why the SM Conditional?

Commenting on the criteria of similarity in Lewis' account, Paul Horwich says:

> Now these criteria of similarity may well engender the right result in each case. However, it seems to me problematic that they have no pre-theoretical plausibility and are derived solely from the need to make certain conditionals come out true and others false. For it is quite mysterious why we should have evolved such a baroque notion of counterfactual dependence. Why did we not, for example, base our concept of counterfactual dependence on our ordinary notion of overall similarity? As long as we lack answers to these questions, it will seem extraordinary that we should have any use for the idea of counterfactual dependence, given Lewis' description of it; and so that account of our conception of the counterfactual conditional must seem psychologically unrealistic. (Horwich 1987: 172)

Horwich's point is that there are infinitely many kinds of conditionals satisfying similarity semantics corresponding to different similarity relations. His question is: 'What is so special about the similarity relation that underlies the conditionals that express counterfactuals?' Neither Lewis nor Bennett addresses this question. I think that the question can be answered for the SM conditional and insofar as Lewis' account (or rather the account repaired to deal with the Albert-Elga objection) approximates the SM conditional account it can be answered for that account too. The answer is, roughly, that the information expressed by SM-counterfactuals is important for us because it tracks the statistical mechanical probability distribution in ways that are important for the consequences of our decisions. Knowledge or partial knowledge of this distribution is relevant to successful decision-making. People whose degrees of belief approximate the statistical mechanical probability distribution are objectively more likely to succeed in satisfying their desires (assuming they are otherwise rational) than people whose degrees of belief diverge from this distribution. So if I know that if I were to strike the match now it is likely light, then I know that P(light/strike&M(now)) is close to 1. If I want to start a fire this knowledge is very useful.

The obvious objection to this proposal is that facts about the statistical mechanical probability distribution are too arcane to be common knowledge. Most people don't know statistical mechanics—the dynamical laws and PROB—and don't know the full Macro state. But if we don't know these, how important can the SM-conditional be? The answer is that we often do know (or believe) many special science generalizations and laws (typically qualified *ceteris paribus*) that enable us to approximate statistical mechanical probabilities. Most of us don't know that this is what we are approximating. So, for example, we know that it is highly likely that the newly fallen snow will melt by evening. On the SM account what we know is made true by the fact that, conditional on the current macro state, the statistical mechanical probability that the newly fallen snow will melt is high. Of course we know only a little bit of the entire macro state—for example that the ambient temperature is above freezing—but most of the rest of the macro state is irrelevant to the melting of snow. The fact that it is irrelevant is, of course, itself a feature of the SM distribution. That our world contains many 'almost isolated' processes is a deep and enormously significant feature of the dynamical laws and the statistical mechanical laws.[45] But because there are such processes and because we know a lot about them we can often know the truth values decision SM-conditionals.

11.5 Conclusion: Future Projects

In this chapter I have continued the project associated with Boltzmann, Reichenbach, Lewis, and most recently Albert of attempting to ground the various arrows of time in statistical mechanics. The main contribution here was to spell out a conditional that approximates the conditional that Bennett and Lewis were attempting to characterize and which we all agree is intimately connected with decision, memory, causation, and so on. The account succeeds in capturing the temporal asymmetry that conditional where Lewis' account fails. Where Bennett puts in the temporal asymmetry by hand this account obtains it from the statistical mechanical distribution; i.e. it is earned not stipulated. Because of the connection between the conditional and the objective statistical mechanical distribution we have the beginnings of an answer to Horwich's question of why conditionals

[45] See Elga (this volume) for the beginnings of an explanation of why this is so.

evaluated in a particular way are especially important. But there is a great deal that is left undone. In the first place the grand scheme of the world composed of PH, PROB and the dynamical laws requires much more defense and development that has been given here. It needs to be investigated exactly how it looks in the context of current accounts of space time and quantum mechanics. Second, Horwich's question deserves a much fuller answer than I gave it here. The SM-conditional is one among many conditionals that could be defined that respect and reflect the probability distribution derived from PROB and PH. More investigation is needed to determine if there is reason to choose this one or any one of these and in some way 'better' for the job of accounting for the other temporally asymmetric notions. Third, the SM-conditional is restricted to conditionals with probabilistic consequents. It remains to be seen if it, or something close to it, can be extended to conditionals whose consequents are not probabilistic since these are closer to ordinary usage. Fourth, I assumed in the paper that the dynamics are deterministic. It remains to be seen if the SM conditional (or a natural extension of it) is appropriate in the context of indeterministic dynamics. Finally, the account will prove its mettle if it can be connected to an account of causation. There is some hope of characterizing causation (or one concept of causation) in terms of counterfactual dependence and there is reason to think that the SM conditional might play the role in a counterfactual analysis along Lewisian lines.

References

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

Beebee, H. (2000). 'The Non-Governing Conception of Laws of Nature', *Philosophy and Phenomenological Research*, Nov. 2000, pp. 571–94.

Bennett, J. (1974). 'Counterfactuals and Possible Worlds', *Canadian Journal of Philosophy*, 4: 381–402.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Callender, C. (2004). 'Measures, Explanation and the Past: Should "Special" Initial Conditions Be Explained?', *British Journal for the Philosophy of Science*, 2004, 55, 195–217.

____ (2003). 'Is There a Puzzle about the Low Entropy Past?', Hitchcock, C. (ed.), *Contemporary Debates in the Philosophy of Science,* chapter 12, Blackwell, 2003.

Elga, A. (2000). 'Statistical Mechanics and the Asymmetry of Counterfactual Dependence,' *Philosophy of Science* suppl. vol 68: 313–24.

Field, H. (2003). 'Causation in a Physical World', in M. Loux and D. Zimmerman (eds.), *Oxford Handbook of Metaphysics*, Oxford: Oxford University Press 2003.

Feynman, R. (1965). *The Character of Physical Law*. Cambridge Mass: MIT Press.

Fine, K. (1975). 'Review of Lewis's Counterfactuals,' *Mind* 84: 451–8.

Horwich, P. (1987). *Asymmetries in Time*. Cambridge Mass: MIT Press.

Kutach, D. (2003). 'The Entropy Theory of Counterfactuals,' *Philosophy of Science*, 69: 82–104.

Lange, M. (forthcoming) 'Laws and Their Stability', *Synthese*.

Leeds, S. (2003). 'Foundations of Statistical Mechanics,' *Philosophy of Science*, 70: 126–44.

Lewis, D. (1986). 'Time's Arrow', in *Philosophical Papers*, vol. II. Oxford: Oxford University Press.

____ (1999). 'New Work for a Theory of Universals', in *Philosophical Papers*, vol. III. Oxford: Oxford University Press.

Loewer, B. (1996). 'Humean Supervenience', *Philosophical Topics*, 24: 101–27.

____ (2001). 'Determinism and Chance', *Studies in the History of Modern Physics*, 32 (4): 609–20.

____ (2003). 'Time and Law' lecture at the Einstein Forum 2003, Berlin.

____ (2004). 'David Lewis' Account of Objective Chance', *Philosophy of Science*, 71: 1115–25.

Maudlin, T. (2005). 'Remarks on the Passing of Time', *Proceedings of the Aristotelian Society*, volume CII (part 3): 237–52 .

North, J. (2004). Time and Probability in an Asymmetric Universe. Dissertation; Rutgers University.

Price, H. (2003). 'There is Still a Problem About the Low Entropy Past' *Contemporary Debates in Philosophy of Science*, ed. C. Hitchcock, Blackwell.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Russell, B. (1913). 'On the Notion of Cause.' *Proceedings of the Aristotelian Society* 13: 1–26.

Sklar, L. (1993). *Physics and Chance*. Cambridge: Cambridge University Press.

Strevens, M. (2003). *Bigger than Chaos*, Cambridge Mass: Harvard University Press.

Tooley, M. (1987). *Causation: A Realist Approach*. Oxford: Clarendon Press.