1

A Modest Proposal Concerning Laws, Counterfactuals, and Explanations

1. METHODOLOGICAL PROLEGOMENON

Philosophical analyses may be pursued via a myriad of methods in service of as great a multitude of goals. Frequently the data upon which an analysis rests, and from which it receives its original inspiration, recount systematic connections between diverse realms of discourse or diverse sets of facts, events, actions, or objects. The aim of the project is elucidating the underlying logical, conceptual, or ontological structure that accounts for these connections. As an obvious example, John's beliefs about what Sarah knows covary systematically with his beliefs about what Sarah believes, about what Sarah has good evidence for, and about what is actually the case. We may explain these covariations by postulating that John at least tacitly adheres to the theory that knowledge is some species of justified true belief.

The results of such a preliminary investigation of correlations among beliefs may be put to various uses. If we choose to endorse John's theory we will simply assert that what Sarah believes, what Sarah has good evidence for, and what is true determine what she knows. We may endorse John's theory, as revealed by his inferences, but criticize his particular judgements. For example, John's inferences may imply that he takes knowledge to require infallible evidence and so, by his own lights, he should not ascribe knowledge to Sarah since her evidence is not conclusive. Or we may instead endorse John's judgements and recommend that he amend his inferences accordingly. And, of course, the inferences, particular judgements, and intuitions at issue may be our own.

This essay was written in 1989, but being too long for a journal and too short for a book only circulated informally. There are evident similarities to John Carroll's approach in his *Laws of Nature* (1994), and we have both been identified as primitivists about laws. I have not attempted a direct comparison between our views as it would not fit into the structure of the paper as originally conceived.

Considerable light may be provided simply by displaying such connections among different realms of discourse. But once embarked upon the voyage, the siren song of reduction is difficult to resist. If the connections can be codified as explicit definitions we can cast out some of the primitive notions with which we are laden, reducing our ontology or ideology. Such reductions may be sought in two quite distinct ways. On the one hand, one may launch the enterprise with some preferred set of concepts, properties, or objects whose philosophical credentials are supposed to be already established. Reduction to this set then validates some otherwise suspect entities. On the other hand, the direction and form of the reduction may await the result of the analysis. Once the inferences that connect various domains have been mapped, one set of items may be found to provide resources to account for some other set. Examples of attempts at the first sort of reduction abound, especially among the logical empiricists. Hans Reichenbach's insistence that purported facts about the geometry of space be parsed into claims about observable coincidences of physical objects plus analytical co-ordinative definitions may serve as a clinical example: the epistemological status of the observable phenomena and of the definitions confer legitimacy upon any further concepts constructed from them (Reichenbach 1958, pp. 10–24). The second approach may be illustrated by David Lewis's realism about possible worlds (Lewis 1986b). Possible worlds hardly recommend themselves on either epistemological or ontological grounds as desirable primitives. Lewis argues, rather, that constructions founded on the plurality of worlds explicate, regiment, and provide semantic grounding for claims about possibility, counterfactuals, propositional content, etc. If we want all these claims to have truth values, Lewis argues, we had best abandon our prejudices and admit possible worlds as primitives in our ontology.

Having sketched a rough taxonomy of philosophical projects, the present endeavor can now be situated. As befits a modest proposal, the direct aims of this enquiry are slight. The primary goal is an outline of the systematic connections between beliefs about laws of nature and a small assortment of other beliefs. The examination will be carried out predominantly in what may be called the conceptual mode, focusing on inferences so as to sidestep the deep problem of ontological commitment. We may liken this to examining John's tacit theory of knowledge without affirming whether or not anyone has any knowledge, whether there is any such thing. As an example of the difference between the conceptual and ontological levels, consider a connection between laws of nature and counterfactuals that has been widely noted: laws, it is said, 'support' counterfactual claims while accidental regularities do not. Such 'support' can be interpreted in two ways. At the conceptual level, it means that if one assents to the proposition that 'all Fs are Gs' is a law, then one will generally also accept that had *s* been an F it would also have been a G. This is a datum about belief formation. On the ontological level, such 'support' would rather represent a relation among objective facts: the law, as a constituent of nature itself, provides grounds for the truth of the counterfactual. One can accept the datum about belief formation but deny any ontological implications by rejecting counterfactuals as candidates for truth. So the result of the investigation may be construed as a conditional: If one wants to assign truth values to counterfactuals then one must also accept laws among the furniture of the world. If one assigns all of the discourse about laws and counterfactuals to the limbo of the non-fact-stating, still the patterns that govern people's willingness to mouth these sentences must be explained.

This enquiry shall not be of the prejudgemental sort. No presuppositions shall be made about the preferability of one sort of discourse to another. Nor should we assume that any reduction must eventuate. If one set of concepts emerges as logically primitive it is because the connections among notions are strong and asymmetrical, allowing some to be generated from others but not vice versa.

Let us begin by setting out the domains of discourse that will be our focus and by sketching their connections to assertions about laws of nature. These connections suffice to upset the most influential philosophical accounts of laws of nature.

2. LAWS, POSSIBILITIES, COUNTERFACTUALS, AND EXPLANATIONS

Beliefs about laws of nature undoubtedly influence and are influenced by any number of other sorts of beliefs. Of these, three classes are of particular interest: beliefs about possibilities, about counterfactuals, and about explanations. Some few examples may illustrate each of these.

A physicist who accepts Einstein's General Theory of Relativity will also believe that it is physically possible for a universe to be closed (to collapse in a Big Crunch) and possible for a universe to be open (continue expanding forever). This is especially evident since we don't yet know whether our own universe is open or closed, so empirical data are still needed to determine which possibility obtains. But even if the issue were settled, the laws of gravitation, as we understand them, admit both possibilities. Anyone who accepts Einstein's laws of gravitation, or Newton's laws of motion and gravitation, must admit the physical possibility of a solar system with six planets even if no such system actually exists. If one believes that the laws of nature governing some sort of event, say a coin flip, are irreducibly probabilistic (and that the outcomes of flips are independent of one another) then one must admit it to be physically possible for any sequence of heads and tails to result from a series of flips.

I take these sorts of inference to be manifest in the actual practice of science and to be intuitively compelling. Any account of the nature of physical laws should account for them.

One connection between laws and counterfactuals has already been noted. If one accepts the conductivity of copper as a law, or as a consequence of laws, then one will also accept, in a wide variety of circumstances, that had a particular piece of copper been subjected to a voltage differential, it would have conducted electricity. Such inferences are notoriously fragile, and in many circumstances counterfactuals seem to have no determinate truth value even though the relevant laws of nature are not contested. This stands in need of explanation. But any acceptable account of laws and of counterfactuals must illuminate the relation of support between the former and the latter where it exists.

Finally, a more amorphous connection is generally acknowledged to hold between laws and explanations. The covering law model, for all its deficiencies, testifies to the depth of this relationship. Coming to see particular events or phenomena as manifestations of laws of nature can provide an understanding of them that does not follow from recognizing them as instances of accidental generalizations. A full elucidation of this fact would require a complete theory of explanation, a task far beyond our scope. But the connection does provide one touchstone for accounts of laws. A law ought to be capable of playing some role in explaining the phenomena that are governed by or are manifestations of it. And a physical event or state or entity which is already explained in all its details by some set of physical laws cannot provide good grounds for appending to these laws new ones. 'We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances' is Newton's first Rule of Reasoning in Philosophy (Newton 1966, p. 398). Any account which disrespects accepted links between laws and explanations thereby loses some of its plausibility.

By way of illustration of these connections, consider a case brought by Bas van Fraassen against the sophisticated regularity account of laws. It is a case to be more fully discussed presently.

To say that we have the concept of a law of nature must imply at least that we can mobilize intuitions to decide on proffered individual examples. Let us then consider a possible world in which all the best true theories, written in an appropriate sort of language, include the statement that all and only spheres are gold. To be concrete, let it be a world whose regularities are correctly described by Newton's mechanics plus law of gravitation, in which there are golden spheres moving in stable orbits about one another, and much smaller iron cubes lying on their surface, and nothing else. If I am now asked whether in that world, all golden objects are spherical because they must be spherical, I answer *No.* First of all it seems to me that there could have been little gold cubes among the iron ones, and secondly, that several of the golden spheres could (given slightly different initial conditions) have collided with each other and thus altered each other's shapes. (1989, pp. 46–7)

The intuitive bite of van Fraassen's example derives from the sorts of connections remarked above. The Newtonian laws of gravitation and motion (plus whatever laws are needed for there to be gold and iron) seem clearly to admit of the possibility of a world such as van Fraassen describes. The shapes and dispositions of the contents of such a world would be set as initial conditions; the stability of the orbits, persistence of the objects, and lack of collisions would then follow from the laws. This scenario might not actually be a physical possibility given that Newton's laws do not obtain in our world. But if we accept that such laws might obtain then the possibility of the laws brings in train the possibility of a concrete situation such as described. The connection between laws and possibilities is manifest.

The connection between laws and explanations also plays a part. Why should we accept that it is *not* a law in this world that something is a sphere just in case it is gold? It can't be, as van Fraassen insinuates, because we already know that different initial conditions would yield non-spherical gold or non-golden spheres. This is a *petitio principii*; if we accept that it is a law, we will not admit the initial conditions as possible. Rather it is because if we assume that Newton's laws are the only laws operating, the sphericality of the gold can be accounted for by initial conditions. Together the initial conditions and Newton's laws entail all of the particular facts about this world. So no new laws need be invoked. This does not prove that there *could not* be a further law about spheres and gold, only that our intuitions accept that there *need not* be. And if there need not be, then a regularity appearing in all the best theories need not be a law.

Van Fraassen's example and variations on it demonstrate that this sophisticated regularity account, associated with Mill, Ramsey, and Lewis, cannot capture the connections between realms of discourse noted above. An acceptable theory of laws must. In the section that follows, I shall argue that none of the main philosophical accounts of laws of nature meets this challenge. Nor can van Fraassen's view, which eschews laws altogether, make sense of actual scientific practice. Our examination will take actual scientific practice as a starting point and return to it in the end.

3. THE LOGICAL FORM OF LAWS OF NATURE

Most philosophical analyses of laws of nature proceed from the logical empiricist formulation of laws of nature, at least as an initial approximation. In that tradition, the logical skeleton of a law is $(x)(Fx \supset Gx)$. A further vague requirement is added that the predicates 'F' and 'G' must be purely qualitative, i.e. contain no ineliminable reference to named individuals. The addendum is added to save the account from total vacuity, for even 'John went to the store' can be pressed into the logical Procrustean bed if the language contains an individual term denoting John and contains identity. But the addendum fails entirely since merely accidental universal concurrences of qualitative properties are possible and since, given a natural stock of spatial relations, each individual can be uniquely denoted by a purely qualitative predicate if the world contains any spatial asymmetry. Despite these drawbacks the logical skeleton serves as a starting point or inspiration for more sophisticated views.

If constraints on the predicates are not sufficient to pick out laws of nature, resort must be made to some more controversial means. One possibility is appeal to modality by the addition of a modal operator: $\Box(x)(Fx \supset Gx)$. The regularity that arises from the operation of laws of nature is neither logical nor metaphysical necessity, so the box must be interpreted as nomic necessity. But appending a box to a sentence and calling it nomic necessity is only marginally superior to appending 'It is a law that …', and can hardly be considered an informative analysis.

Perhaps the difficulty lies instead in the horseshoe. Material implication is notoriously weak, so some more metaphysically substantive connective may be of use. It is not just that $(x)(Fx \supset Gx)$ is necessary, it is that being an F necessitates being a G. According to the view developed by David Armstrong,[1] a law holds if being an F necessitates being a G in virtue of a relation between

[1] Similar views have been developed by Micheal Tooley (1977) and Fred Dretske (1977). I am using Armstrong as an exemplar.

the universals F-ness and G-ness. In Armstrong's formulation the law is symbolized as N(F,G)(a's being F, a's being G), from which the offending horseshoe has been eliminated (Armstrong 1983, p. 90). This elimination, though, immediately poses a new problem, for $(x)(Fx \supset Gx)$ is supposed to follow from the law. In the view of some, e.g. van Fraassen, the gap between a fact about relations of universals and a fact about relations among their instances cannot be closed (van Fraassen 1989, pp. 97 ff). Further, Armstrong's approach may be considered *ignotum per ignotus*: do we really have a firmer grasp of what a necessitation relation between universals is than we do of what a law is? This is especially troubling since the species of necessitation at issue must again be denominated *nomic* necessitation.

Armstrong's account, although eschewing $(x)(Fx \supset Gx)$ as any sort of law, is still naturally regarded as influenced by that skeleton. The formula directs our attention to a relation between the predicates 'F' and 'G', and the universality of laws, enshrined in the quantifier, implies the universality of the relation. It is not a far step to suppose that this universal relation among instances derives from a relation among the universals denoted by the predicates. Universality was the only feature of laws that the positivists could get a clean syntactic grip on, and it continues to influence even those who firmly reject the positivist view. Indeed, Armstrong is able to admit the possibility of purely local laws only by the dubious admission of universals that are not purely qualitative (Armstrong 1983, p. 100).

Given that the preliminary supposition about the logical form of laws can so strongly influence subsequent avenues of research, we ought to pause to ask whether $(x)(Fx \supset Gx)$ is really a useful place to begin. Does it contain features that laws lack? Does it ignore structure commonly found in laws? An appropriate place to begin is with some real scientific theories rather than with cooked examples of the 'All ravens are black' variety. Let us look at some laws without formal prejudices and see what we find.

The fundamental law of Newtonian mechanics, the mathematical consequence of Newton's first two laws, is $\mathbf{F} = m\mathbf{a}$ or $\mathbf{F} = m\, d^2\mathbf{x}/dt^2$ or, most precisely, $\mathbf{F} = m\, d(m\mathbf{v})/dt$. The fundamental law of non-relativistic quantum mechanics, Schrödinger's equation, is $i\hbar\, \partial/\partial t\, |\Psi\rangle = H\, |\Psi\rangle$. No doubt these can be tortured into a form similar to $(x)(Fx \supset Gx)$, but it is hard to see what the purpose of the exercise would be. What is most obvious about these laws is that they describe how the physical state of a system or particle evolves through time. The laws are generally presumed to be universally valid. But this is not a feature directly represented in the formulae, nor does it appear to be essential to their status as laws. It is not contradictory to assert, or does not at first glance seem to be, that the evolution of physical states of particles is governed by Newton's laws *around here*, or that it has been *for the last 10 billion years* (but not before that). John Wheeler has proposed that after the Big Crunch the universe will be 'reprocessed' probabilistically, with the constants of nature and of motion, the number of particles and mass of the universe being changed (Misner, Thorne, and Wheeler 1973, p. 1214). It is a small step to suggest that the very laws themselves may change. Nor does it sound absurd to suggest that the laws elsewhere in the universe may differ from those here, especially if the various regions are prevented from interacting by domain walls of some sort. There might be some meta-law governing these different laws, or there might not be. But the supposition that Schrödinger's equation describes the evolution of physical quantities only in this 'bounce' of the universe, between our Big Bang and Big Crunch, doesn't seem incompatible with describing it still as a law, albeit a parochial one. At least, such a discovery would not appreciably alter our assessment of possibilities, counterfactuals, and explanations in most contexts. Astrophysics, biology, chemistry, physiology, and everyday language would be unchanged by the discovery. Parochial laws are still laws.

The laws cited above, then, tell us how, at least for some period and in some region, physical states evolve through time. Standing alone they are incomplete, for we need principles for determining the forces in Newtonian mechanics and the Hamiltonian operator H in quantum theory. Newton's third law and law of gravitation supply part of this demand. But the principle of temporal change is the motor of the enterprise. Supplying a force function for electrical interactions, frictional forces, etc. yields instances of Newtonian mechanics. One can change the form of a force function but stay within the Newtonian regime. Changing the law of temporal evolution, though, constitutes a rejection of Newtonian mechanics. Similarly, Schrödinger's equation, without any further specification of the Hamiltonian operator, is considered a fundamental principle of quantum mechanics. The specification of the Hamiltonian is a task that changes from one physical context to another.

Let us call a proposed basic law that describes the evolution of physical states through time a Fundamental Law of Temporal Evolution (FLOTE). Other sciences and folk wisdom recognize generalizations about temporal development that may be regarded as more or less lawlike, such as the Hardy–Weinberg law of population genetics or the observation that certain plants grow towards sources of light. Let us denominate these simply Laws of Temporal Evolution (LOTEs), being lawlike insofar as they are accepted as supporting counterfactuals and as supplying explanations. LOTEs are happily acknowledged to admit of exceptions (e.g. if a nuclear explosion occurs nearby, the plant won't grow towards light sources). But they still are accepted as describing how things would go, at least approximately, under normal conditions. LOTEs are generally thought to be ontologically parasitic on FLOTEs: in our world the laws of population genetics describe temporal changes in the gene pool in part because the laws of physics allow the physical realizations of the genes to have the properties—such as persistence through time, ability to recombine, etc.—which population genetics postulates of them. There is no such inverse dependence of FLOTEs on LOTEs.

Beside FLOTEs there are the adjunct principles that are needed to fill out the FLOTEs in particular contexts, principles about the magnitudes of forces and the form of the Hamiltonian, or about the sorts of physical states that are allowable. Some of these, such as Newton's law of gravitation, are laws of coexistence; others, such as the superselection rules of quantum mechanics, are constraints that are not naturally construed as either laws of succession or of coexistence. Some so-called laws of coexistence, such as the ideal gas law $PV = nRT$, are better construed as consequences of laws of temporal evolution. $PV = nRT$ is true only of equilibrium states: if we rapidly increase the volume of container, the pressure of the gas inside ceases to be well defined until after the relaxation time characteristic of the system has elapsed. The gas evolves into a state satisfying the law, and remains in such a state only so long as it is in equilibrium. Sorting out the status of these various adjunct principles and consequences of law is a task requiring nice discriminations that is foreign to our present purpose.

What is clear is that scientific and commonsense explanations demand the postulation of (F)LOTEs and their adjunct principles. It is only barely possible to conceive of a world that displays no diachronic regularity at all, in which earlier states are not even probabilistically associated with later ones. No object could persist, so the world would consist in point events forming no pattern through time. In a Newtonian setting, there still might be laws of coexistence among these events; in a relativistic regime, where there is no preferred simultaneity relation, total chaos would reign. And this is on the generous assumption that the very notion of a spatio-temporal structure could be defined absent any laws of temporal evolution.

FLOTEs can be deterministic or irreducibly stochastic. If the former, then the specification of appropriate boundary conditions suffices to fix the physical state in an entire region; if the latter, boundary conditions together with the law only determine a collection of possible physical states with associated probabilities.

To sum up this section, a look at laws used in science reveals a basic sort of law, the Law of Temporal Evolution, that specifies how specified states of a system will or can evolve into other states. Other laws are adjuncts to these, having content only in association with a FLOTE or with a LOTE designed for a more abstract level of description of the physical state. Such LOTEs apply to our world in virtue of the operation of FLOTEs. These laws may or may not be universal; in principle they might govern only a limited region. In the remainder of this paper we will consider how the idea of a FLOTE and its adjunct principles can illuminate the connections among laws, possibilities, counterfactuals, and explanations.

4. THE MODEST PROPOSAL

We have provided a rough characterization of laws of temporal evolution and adjunct principles. Taken together they provide descriptions of the state of a system and a rule, deterministic or probabilistic, of how that state evolves through time. So far no sort of philosophical analysis of these laws, or of lawhood, has been advanced. The temporal career of the world displays all sorts of regularities, described by more or less complex mathematical functions, including the accidental regularities that brought the logical empiricist account to grief. To make things worse, I have defended the claim that the regularities due to law need not persist through all time or obtain in all places since the laws may not. So what makes some such regularities into laws of nature, or into the consequences of law?

This question can take two forms. On the ontological side we may seek some further fact or structure that confers lawhood. For Armstrong, to give an example, the relevant further structure is a relation between universals. On the epistemological side, we may ask how we know which observed regularities are consequences of laws as opposed to being accidental. Armstrong must adopt at least some degree of skepticism here since no *observable* relations among instances of universals guarantee that there exists the relation of necessitation among the universals themselves. Even an ideal observer who sees everything that can be seen in the whole history of the universe cannot be entirely confident to have gotten the laws right.

On Lewis's sophisticated regularity view[2] the two questions have a single solution. What makes a regularity into a law is that it appears in all of the simplest, most informative true theories of the world (Lewis 1973a, pp. 72–7, see also 1986a, pp. 122–31). We can get at least presumptive evidence about what the laws are by formulating as many such theories as possible; an ideal epistemic agent provided with all of the particular facts about a world could, in principle, determine the laws from this information.

My own proposal is simple: laws of nature ought to be accepted as ontologically primitive.[3] We may use metaphors to fire the imagination: among the regularities of temporal evolution, some, such as perhaps that described by Schrödinger's equation, govern or determine or generate the evolution. But these metaphors are not offered as analyses. In fact it is relatively clear what is asserted when a functional relation is said to be a law. Laws are the patterns that nature respects; to say what is physically possible is to say what the constraint of those patterns allows.

Taking laws as primitive may appear to be simple surrender in the face of a philosophical puzzle. But every account must have primitives. The account must be judged on the clarity of the inferences that the primitives warrant and on the degree of systematization they reveal among our pre-analytic inferences. Laws are preferable in point of familiarity to such primitives as necessitation relations among universals. And if we begin by postulating that at each time and place the temporal evolution of the world is governed by certain principles our convictions about possibilities, counterfactuals, and explanations can be regimented and explained.

As an example, let us return to van Fraassen's example and variations on it. The appeal of Lewis's account is that it requires no additions to our ontology beyond particular matters of fact. Events occur in space and time, and if these events form patterns we do not swell our ontology by recognizing them. If we single out some of these regularities by somewhat pragmatic considerations,

[2] This view is also associated with Ramsey and with Mill. I take Lewis as exemplar because he has developed the most detailed account.

[3] To give this a conceptual level reading: the idea of a law of nature is not logically derived from, and cannot be defined in terms of, other notions.

by how they cohere to form simple and informative theories of the world, we do not add to the furniture of the world. But this account of laws fails to accord with our beliefs.

It is possible for a world governed solely by the laws that govern our world to be such that every person who reads *The Satanic Verses* is subsequently crushed by a meteor. It would take a massive coincidence, but such a coincidence will result from a possible set of initial conditions: meteors could exist on just the right trajectories. If cosmic rays were distributed in just the right way, every person who drinks milk might subsequently contract cancer. The examples could be multiplied indefinitely. The models of quantum mechanics and of Newtonian mechanics can display, due to special initial conditions or the fortuitous outcomes of random processes, accidental regularities as widespread and as striking as one pleases.

According to Lewis's view, in such a world, if the regularities are striking enough, new laws exist. The criterion for selection of laws on his view is vague, requiring a balance of simplicity and informativeness. But any such standard can be met by accidental regularities. 'All stars are members of binary star systems' is simple and highly informative, and there are models of Newton's theory of gravitation in which it is true. But for all that, we do not regard it as a law in those models: it is a consequence of special initial conditions. The Lewis view cannot admit this as a possibility, but to any astrophysicist or cosmologist it plainly is. The determination whether the universe is opened or closed will be the discovery of a simple, highly informative fact about the universe, but it will not be the discovery of a law. The initial conditions leading to a closed universe are possible, as are those leading to an open one. Whether the universe is open or closed is determined by its initial conditions and by the laws of gravitation. The fact is neither a law nor a consequence of laws taken alone.

The point is even more acute for stochastic systems. If coin flips are governed by irreducibly probabilistic laws then all sequences of results, including all heads or all tails, are physically possible. And in a world in which all of the flips happen to come up heads the proposition that they do so is both simple and highly informative. The inhabitants of such a world would doubtless take it to be a law that flipped coins come up heads, and they would be rational to do so, but they would be victims of a run of bad luck. If only finitely many coins are flipped we can even calculate the likelihood, given the probabilistic laws, of such an unfortunate event. If the number of flips is infinite the likelihood is zero, but so is that of any particular sequence. All heads is still possible in a world governed by such a law.

It is inconsistent to claim that while a law obtains the world can evolve in such a way that the law fails. If a probabilistic law governs coins then the world may evolve such that 'All coin flips yield heads' is part of the simplest, most informative theory of that world. If 'All coin flips yield heads' were a law then coins could not be governed by a probabilistic law. Hence being a member of that simplest, most informative theory cannot be sufficient for being a law.

To the ontological question of what makes a regularity into a law of nature I answer that lawhood is a primitive status. Nothing further, neither relations among universals nor role in a theory, promotes a regularity into a law. FLOTEs, along with their adjunct principles, describe how states may evolve into later states. If a law governs a particular space-time region then the physical states there will so evolve.

To the epistemological questions I must, with Armstrong, admit a degree of skepticism. There is no guarantee that the observable phenomena will lead us to correctly infer the laws of nature. We may, for example, be the unlucky inhabitants of an 'all heads' world governed by a stochastic law. We may inhabit a universe whose initial conditions give rise to regularities that we mistakenly ascribe to laws (we will see a very concrete example of this possibility in section 7). The observed correlation between distant events so important to the confirmation of quantum theory may be a run of bad luck: they could be the results of independent stochastic processes that accidentally turn out to be correlated. If so, then some of our best evidence is systematically misleading, and in rationally assessing the evidence as supporting the theory we are wandering away from the truth.

Laws are ontological primitives at least in that two worlds could differ in their laws but not in any observable respect. The 'all heads' probabilistic world looks just like another possible world that is governed by a deterministic law. If our ontology includes objective single-case propensities the two worlds will differ in their propensities. But if we wish to make laws ontologically derivative in that they supervene on the global distribution of non-nomic entities, then simply admitting objective single-case propensities will not do the job. For the probabilistic world in which these *propensities* are fixed by a deterministic law will not differ (at least in first-order propensities) from a world in which the single-case propensities are governed by a stochastic law and in which all the coin flips, beside just happening to come up heads, also just happen to get a 50 per cent propensity for coming up heads. We can introduce second-order propensities for the first-order ones, but with the obvious rejoinder. Better to regard the stochastic laws as absolutely ontologically primitive and explain single-case propensities as consequences of falling under such laws.

My analysis of laws is no analysis at all. Rather I suggest we accept laws as fundamental entities in our ontology. Or, speaking at the conceptual level, the notion of a law cannot be reduced to other more primitive notions. The only hope of justifying this approach is to show that having accepted laws as building blocks we can explain how our beliefs about laws determine our beliefs in other domains. Such results come in profusion.

The first obvious connection is to physical possibility. Our world seems to be governed by laws, at least around here. When we say that an event or situation is physically possible we mean that its occurrence is consistent with the constraints that derive from the laws. The possible worlds consistent with a set of laws are described by the models of a theory that formulates those laws.

If the laws are continuous and deterministic then the models are easily characterized. For simplicity, let us take a deterministic FLOTE and adjunct principles that operate in a special relativistic space-time. Take a surface that cuts every maximal timelike trajectory in the space-time exactly once (a Cauchy surface). Specifying a Cauchy surface is the analog to choosing a moment of time in a Newtonian regime; roughly one can think of it as a surface that cuts the space-time in two horizontally (the vertical direction being timelike). Boundary values can be specified on this surface, such as the distribution of particles, intensities of fields, etc. In some cases the data are freely specifiable, in some (due to adjunct principles) they are subject to constraints. In either case there is a well-defined class of admissible boundary values. The FLOTE now specifies how those values will evolve through time. If the FLOTE is deterministic in both the past and future directions, then the boundary values will determine a unique distribution of the physical magnitudes through all time. Such a distribution describes a physically possible world relative to those laws.[4]

If the FLOTE is stochastic then the situation is messier but still pretty clear. Specific boundary values on the Cauchy surface yield not a single model

[4] Lots of corners are being cut here. If the space-time is Newtonian and no constraints are put on maximum velocities, then the Cauchy surface must include a surface which surrounds a given system through all time, with 'incoming' data specified. If the laws include the General Theory of Relativity then the space-time itself is generated, not fixed. The mathematical and physics literature on boundary value problems is vast, and John Earman's (1986) is a wonderful guide to some of the intricacies, but these details do not affect our general picture. The point is that the class of models of a theory is well defined and is isomorphic to the possible worlds allowed by the laws described by the theory. Further, data on a small portion of a world together with the laws can determine data throughout the world, or throughout a large region.

but a set of models, corresponding to all of the outcomes permitted by the laws. Furthermore, the set of models consistent with some boundary values is invested with a metric over measurable subsets, a measure of how likely it is, given those boundary values, that the world will evolve into one of the members of that subset. If the boundary conditions specify that 100 coins are about to be flipped then the set of models associated with the probabilistic law of unbiased coins (50 per cent likelihood of heads) and with a law of biased coins (e.g. 80 per cent likelihood of heads) are identical: they each contain a model for each possible combination of heads and tails. But the probabilities assigned to these models by the two theories differ.[5]

Given a FLOTE and adjunct principles, then, the notion of physical possibility relative to those laws can be immediately defined. And the result is intuitively correct. Newtonian mechanics allows for a world in which all spheres are gold but it is not a law that all spheres are gold. Stochastic laws allow for the possibility of uniform results that are not a matter of law. The objections to the Lewis view are avoided and the intuitions that backed the objections are explained.

Above it was suggested that the notion of a stochastic law should be taken as more primitive than that of an objective single-case propensity. Propensities can now be easily derived. The propensity for some occurrence, given a set of boundary conditions on a Cauchy surface, is just the probability assigned by the stochastic laws to the set of models with those boundary conditions and that outcome. Given some actual model, it may be the case that the propensity assigned to an occurrence relative to any Cauchy surface that cuts within some temporal distance D before the occurrence is the same, or that the propensities assigned by surfaces that cut within D approach a limit as D approaches zero. If so, then that limit is the objective single-case propensity for the event.

As a small bonus, we can also see what is so peculiar about the distant correlations of quantum mechanics. Take a model of a stochastic theory defined on a relativistic space-time. Consider an event *e*. Take any two Cauchy surfaces that overlap where they intersect the back light-cone of *e*, although they may differ outside the back light-cone. We will say that *e* is the result of

[5] The official definition assigns the probability to subsets of the set of models rather than to individual models because for stochastic processes with a continuous range of possible outcomes (e.g. radioactive decay, which can occur at any time), or models with an infinite number of discrete stochastic events, the probability of each model may be zero. Still, specified subsets, such as those in which the atom decays within a particular time period or those in which a particular discrete event has a given outcome, may be assigned a definite probability. Of course, individual models may be assigned non-zero probabilities by certain theories.

a *local stochastic process* if the probability assigned to *e* by the theory relative to any such pair of Cauchy surfaces is the same. The notion of a local stochastic process is slightly weaker than that of having an objective propensity. If we add that the probabilities assigned by all continuous sequences of Cauchy surfaces should approach the same value as the maximum timelike interval between the event and the intersection of the Cauchy surfaces with the back light-cone approaches zero, then any result of a local stochastic process will have an objective probability. That is, in such a case not only will variations in a Cauchy surface outside the back light-cone not affect the probabilities, also how the surface cuts across the back light-cone does not affect them.

The motivation of the definition is straightforward: according to (at least one interpretation of) Relativity, events should be influenced only by other events in their back light-cone. So Cauchy surfaces that agree on the back light-cone and have the same boundary values there should agree on everything that could possibly be relevant to the occurrence of the event. Therefore the theory should assign the same probability to the event on the basis of data on any such Cauchy surface.

The problem with quantum mechanics is now quickly stated: the probabilities for correlations between distant events assigned by quantum mechanics cannot be reproduced by any theory in which all events are the result of local stochastic processes. Quantum mechanics cannot be reconciled with the interpretation of Relativity stated above.

FLOTEs give us immediate purchase on the notions of physical possibility and objective propensity. And once we have the set of models, physical necessity is easily defined in the usual way, using the models as mutually accessible possible worlds. So in a way, the problematic inherited from Hume has been solved by turning the direction of analysis around. Here is what I mean.

Hume begins by investigating the notion of cause and effect, and finds within it a notion of necessary connection between events. He then worries about giving an empiricist account of the origin (and hence content!) of the notion of this necessary connection, and finds that he is led either to constant conjunction between events or to a subjective sensation that accompanies an inference bred of long experience of constant conjunction. The 'necessity' must reduce either to mere pattern or to a purely subjective sensation, and in neither case pertains solely to the two events thought to be necessarily conjoined. Although Hume does not focus so intently on the motion of a law of nature, the natural implication is that laws can be nothing but patterns of events either.

I take content of the laws to be expressed by equations like Newton's equations of motion, and the status of lawhood to be primitive. What then of the notion of 'necessary connection'? The content of the laws can be expressed without modal notions, and suffices to determine a class of models. The models can then be treated as 'possible worlds' in the usual way, and so provide truth conditions for claims about nomic possibility and necessity. The laws themselves, of course, turn out to be nomically necessary, since they obtain in all the models. We can give a clear account of the 'had to' in claims like 'The bowling ball had to fall once the support beneath it was removed': in every model of the laws with an unsupported bowling ball, the bowling ball falls. So we have all the 'physical necessity' we need without having invoked anything beside the laws. And given the laws we can easily construct truth conditions for even more.

5. FLOTES AND COUNTERFACTUALS

One who regards laws as primitive will also regard them as quite definite. At a given time the future temporal behavior of the world is constrained in some exact way by the laws, irrespective of our beliefs, desires, and concerns, irrespective of pragmatic considerations or contexts. On the other hand, the evaluation of counterfactual claims is widely recognized as being influenced by context and interest. This contrast poses a challenge for those who seek to explicate the bearing of laws and counterfactuals on one another. The challenge is more difficult for anyone who holds that our judgements about laws depend derivatively on previously settled beliefs about counterfactuals: it is not easy to create a rigid structure if some of the basic components are elastic. In the opposite direction thing go more smoothly, for if judgements about counterfactuals depend on beliefs about laws and on other things, and if the other things reflect pragmatic considerations, then judgements about the counterfactuals may be variable or indefinite although the beliefs about laws remain fixed.

What follows is not a unified theory of all counterfactuals. It is likely that no such theory exists. What would have happened if it had been 20° warmer at Canaveral on the day the *Challenger* exploded? If trees could walk? If Plato and Carnap had been contemporaries? If gravity were an inverse cube force? If one could trisect the angle with ruler and compass? We are wont to consider what would have happened if physical conditions had been different, if laws were changed, if metaphysical necessities were violated, even if logical truths failed to hold. The methods of evaluation of these claims are bound to be diverse. Even Lewis's theory of counterfactuals must give out in the last case: it is non-trivially true that if one could trisect the angle one could square the circle. Lewis can get truth, but not non-triviality: his theory cannot explain why a mathematician would try to prove the claim (Lewis 1973a, pp. 24, 25). And Lewis can certainly not explain discriminations of such counterfactuals as true *or* false, e.g. it may be false that if one could construct a 23-gon one could square the circle. It is bootless to rest the semantics of counterfactuals on relations to possible worlds in this case.

Still, a large number of types of counterfactuals seem to be treated similarly, and, more important, those so treated are among those for which our intuitions are strongest. We will begin with cases that are as uncontentious as possible and are closely tied to physical law. If we understand what makes these cases uncontroversial we will be able to predict under what circumstances doubts about the truth value or the meaningfulness of counterfactuals will begin to creep in.

If the bomb dropped on Hiroshima had contained titanium instead of uranium it would not have exploded. If Lindbergh had set out with half as much fuel, he would not have made it across the Atlantic. If the ozone layer in the atmosphere should be destroyed, the incidence of cancer will increase. These claims seem undoubtedly true. In the agora of everyday discourse and in scientific contexts such claims are treated on a par with descriptions of the *Spirit of St Louis* and assessments of the chemical composition of the atmosphere. Even if they are ultimately shown to be counterfeit currency, passing as statements of fact when they play some other role, one must account for the assurance with which such claims are made and evaluated. (In defense of their legitimacy it is notable that the last, the future subjunctive, may or may not be counterfactual for all we know now. That doesn't affect the means we use to verify it.)

I take the three sentences above to be true, and I take their truth to depend on the laws of nature. They also depend on other factors. Let us start with the first case.

We have already seen how FLOTEs plus boundary conditions on a Cauchy surface generate models. For simplicity, suppose that the laws governing our world are local and deterministic, as they essentially are for the case at hand. Extension to stochastic laws will come later.

We wish to know what would have happened if the bomb had contained titanium in place of uranium. Here is the recipe. Step 1: choose a Cauchy surface that cuts through the actual world and that intersects the bomb about the time it was released from the plane. All physical magnitudes take some value on this surface. Step 2: construct a Cauchy surface just like the one in Step 1 save that the physical magnitudes are changed in this way: uranium is replaced with titanium in the bomb. Step 3: allow the laws to operate on this Cauchy surface with the new boundary values generating a new model. In that model, the bomb does not explode. Ergo (if we have got the laws right, etc.) the counterfactual is true.

In this three-step process laws come into play essentially, but only at the last stage. If we manage to select a unique Cauchy surface and to alter its data in a unique way, and if the laws are deterministic, then all counterfactuals with that antecedent will have determinate truth values. For a single model will thereby be specified and the consequent of the counterfactual will either be true in it or not.

Even if there is some ambiguity in the Cauchy surface and in the way to change the boundary values, still the claim may have a determinate truth value. Because of the ambiguity many different surfaces may be selected and the data on them changed in many ways. The result is a set of models rather than one: a model for each acceptable changed surface. If the consequent obtains in all of the models, the counterfactual is true; if in none, false. If it obtains in some but not others, the counterfactual has an indeterminate truth value and our intuitions should start to get fuzzy. Examples will bear these predictions out.

The purpose of the antecedent of a counterfactual is to provide instructions on how to locate a Cauchy surface (how to pick a moment in time) and how to generate an altered description of the physical state at that moment. The antecedent is not an indicative sentence at all; it is more like a command. If the command is carried out a depiction of a situation will result and according to that depiction a certain indicative sentence will be true, but the command is not the same as the indicative sentence. Despite surface similarities to the material conditional, the counterfactual conditional is not a two-place function whose arguments are two propositions. It is a function whose first argument is a command and second is a proposition. This will explain why the counterfactual conditional fails to have any of the formal features of the material conditional: transitivity, contraposition, and strengthening of the antecedent (cf. Lewis 1973a, pp. 31 ff.).

Thus the 'if the bomb dropped on Hiroshima had contained titanium instead of uranium …' directs us to a moment shortly before the atomic explosion and instructs us to alter the description of the state of the world so that titanium replaces uranium in the bomb. And there is a tacit ceteris paribus condition: leave everything else the same. Don't fool with the position of the plane or the wiring of the bomb or the monetary system of France or anything else. Similarly, if I command you to take the roast beef in to the guests you have not carried out the command if you step on the roast beef first, and if you murder one of the guests in the process you did not do so on my instructions. Of course, one cannot just change the uranium into titanium and leave *everything* else the same. The total amount of titanium in the universe will be changed, as will the ratio of titanium to steel in the bomb. So the clarity of the instruction, what counts as fulfilling it and what not, depends on the clarity of the ceteris paribus clause, and this is a function of the context.

The ideal case obtains when the physical state of the world is understood as specified by providing values of a set of independent variables and the command is to change one of those variables. Then 'ceteris paribus' means 'leave the rest of the variables alone'. In Newtonian mechanics, such variables could be the positions, masses, and velocities of all the particles. An instruction to change the mass or velocity of a certain particle will be, in this context, unambiguous. But such an analysis of the physical state into a set of independent variables is not provided by the laws or by Newtonian theory as a whole. We can specify states by giving masses, positions, and velocities, or equally well by masses, positions, and momenta. If I am told to increase the mass of a particle this difference may have an effect: increasing the mass while leaving the velocity unchanged increases the momentum; increasing the mass while leaving the momentum unchanged decreases the velocity. In each case different *cetera* are *paria*, and which change is appropriate is decided, if at all, by context and background assumptions.

The independent variables might, in principle, be quite bizarre. Instead of independently specifying the masses of each particle, we could give the mass of a chosen particle and the mass ratios of the rest to the standard. In this context, increasing the mass of the standard particle and leaving everything else the same entails increasing the masses of all the others. Of course, a counterfactual instruction would have to warn us if this were meant, for the choice of variables is highly unusual.

In most contexts we have a rough and ready sense of which variables are to be treated as independent of one another. But the instructions may be vague in other ways. 'If Lindbergh had set out with half as much fuel …' tells us to choose a moment near Lindbergh's departure and change the state of the world so his fuel tank contains half as much gasoline. Don't change his hair color or the structural members of the *Spirit of St Louis* or the height of the Eiffel tower. But we are to delete the extra fuel and replace it with ... what? Not a perfect vacuum. Not water. Air, presumably. But the instruction doesn't tell us to replace it with air, so this information must come from background assumptions. When such background assumptions do not suffice to clarify the means of carrying out the command, the counterfactual may fail to be determinate.

Just as the instructions for Step 2 may be more or less vague, so may the instructions for Step 1, choosing the moment of time. Commonly this ambiguity will make no difference to the outcome, sometimes it will. Consider some data. What are the truth values of the following counterfactuals? Case 1. Tom is standing in a field two meters north of Sue. A large meteor crashes down, incinerating Tom. Counterfactual A: If Sue had been standing two meters north, she would have been killed. Case 2. Tom, a famous politician, is standing two meters north of Sue. A crack sniper, who has been stalking Tom, fires and assassinates him. Counterfactual B: If Sue had been standing two meters to the north, she would have been killed. Counterfactual C: If the trajectory of the bullet had taken it two meters south, Sue would have been killed. Counterfactual D: If the bullet had been deflected two meters south, Sue would have been killed.

The reader is requested to consult her or his intuitions on these cases. The author and several colleagues questioned are in broad agreement. Counterfactual A is clearly true. Counterfactual B is hazy, a common comment being 'needs more context'. Counterfactual C is probably true, although this is not so clear as case A. D is clearly true.

The recipe delivers these results. Return to any moment shortly before the meteor strikes and move Sue two meters north, in any reasonable way. Dump Tom somewhere else. The trajectory of the meteor is unchanged and the laws generate a situation in which Sue is hit. In the second case, though, the exact moment chosen is of great importance and different choices give different outcomes. If a moment is selected after the shot is fired but before it hits Tom, Sue will be shot. If the moment is somewhat before the shot was fired, the natural laws yield a different outcome. The sniper will notice that he is not aiming at Tom and will compensate. Sue will be safe. The third case has both the ambiguity of time and of the instruction. How should we alter the trajectory? If by altering the aim of the gun then we have two cases. We can keep Sue's position unchanged, and she will be hit. Or we can keep the expertise of the marksman unchanged, in which case he would only be aiming where Tom is, so Tom must be moved. The expertise of the sniper is especially important because it is one of the few feature of the situation explicitly stated. So C can get different outcomes. D is entirely clear: the bullet is to be changed en route, so the expertise of the marksman is not in question. All models yield the same result.

An instruction or command is not a proposition, nor can it be construed as an attitude adopted toward a proposition. At least this is so if one considers 'A weighs the same as B', 'B weighs the same as A', and 'A and B weigh the same' to express the same proposition. For instructions command not only that one change a state so as to make a certain proposition true, they also indicate how the change is to be effected. Roughly, 'If S were P ...' instructs us to change S so that it is P. Examples: (1) If Laurel weighed the same as Hardy, the pair would have weighed over 500 pounds. (2) If Hardy had weighed the same as Laurel, the pair would have weighed under 400 pounds. (3) If Laurel and Hardy had weighed the same ...? (In these cases, evaluating the counterfactual requires no third step since the truth value of the consequent is already determined at Step 2.) Instruction 1 tells us to change Laurel to make the proposition true, instruction 2 to change Hardy. Instruction 3 is unclear. To evaluate counterfactual 2 by fattening Laurel up at Step 2 would be as perverse, and as incorrect, as a weight expert who set about fulfilling the command to make Hardy weigh the same as Laurel by feeding Laurel.[6]

Different contexts can demand that counterfactuals be evaluated by methods that accord to a greater or lesser degree with the recipe. Those more in accord with it give the more correct counterfactuals about what would actually happen. Examples: (a) The ant can lift 500 times its body weight. So an ant the size of a human could lift a VW bus. (b) Weight increases as the cube of linear dimension. Tensile and compressive strength of structural materials increases as their cross-section, i.e. as the square of linear dimension. Hence an ant the size of a human would collapse under its own weight (see Galilei 1974, second day). The first counterfactual stipulates (by context) the factor to remain unchanged at Step 2: weight-to-lift ratio. The consequence follows by logic after Step 2, Step 3 is unneeded. The second scales up the creature at Step 2 and then lets the laws governing the materials determine the outcome. It is the correct result if we are concerned with physical possibility. A small model of a skyscraper made of paper may stand up perfectly well.

[6] This problem, as well as most of the others discussed here, was pointed out by Nelson Goodman (Goodman 1983, chapter 1).

A scaled-up full-size model made of scaled-up paper would collapse. That would also be the fate of a gargantuan ant.

We can now explain the failure of counterfactuals to display the formal features of the material conditional. Failure under strengthening of the antecedent is manifest. If Lindbergh had had half as much fuel he would not have made it. If he had had half as much fuel and an engine twice as efficient he would have. The first instructs us to reduce the fuel and leave the engine alone, the second to change both fuel and engine. It is hardly surprising that the results of each instruction might differ radically.

Failure of transitivity is equally manifest. If A had occurred, B would have occurred. But the *way* B would have occurred might be very different than the way we would bring it about if instructed to. If the Earth had exploded in 1987, Ivana would not have found out about Marla. If Ivana had not found out about Marla, Trump would be a happier man now. In normal contexts both of these are true. It does not follow that if the Earth had exploded, Trump would be happier. The instruction to go back and prevent Ivana from finding out is vague in the extreme; one could bring it about by innumerable different mechanisms, but apocalyptic cataclysm is not among them.

There is a context in which transitivity holds, which is when we are *continuing* a scenario. If A had occurred, B would, and if B had occurred *in that way* C would. So if A had occurred C would. Because of this, we may feel a bit queasy about evaluating the second counterfactual above given its proximity to the first, even though in neutral contexts we would readily assent to it.

Contraposition fails for the same reasons. If the antecedent and consequent describe events at different times then one of the counterfactuals must take the more unusual form 'If A had happened B would have had to have happened.' Consider an undecided voter who decided at the last moment, and with little conviction, to vote for Reagan. Several years later, unhappy with the state of the country, she consoles herself: 'Well, even if I hadn't voted for him he still would have been elected.' This is true. The contraposition, 'If Reagan had not been elected, then she would have had to have voted for him,' is certainly not true, and one could make a case that it is false.

It is also obvious that contraposition must fail on purely syntactic grounds. The consequent of a counterfactual is a proposition, so the truth value of a counterfactual is not changed under substitution in the consequent of sentences that express the same proposition. We have already seen that this is untrue for the antecedent. Under contraposition, the sentences in the antecedent and consequent change roles, so changes that can make no difference for the original counterfactual can change the truth value of the contraposition. If contraposition always held, we could counterpose 'If Laurel weighed the same as Hardy' in the antecedent to 'Laurel would not have weighed the same as Hardy' in the consequent, switch the consequent to 'Hardy would not have weighed the same as Laurel', then counterpose this back to 'If Hardy weighed the same as Laurel' in the antecedent, all without changing the truth value. But we have seen that this is impossible.

So far we have considered only cases where all of the ambiguities and vaguenesses result from the instructions on how to carry out Steps 1 and 2 of the recipe. At Step 2 we have a set of Cauchy surfaces with changed boundary values, which the laws of Step 3 extend forward (or backward) in time to yield a model. Rather deep difficulties appear when we attempt to extend this account to stochastic laws. These difficulties are accompanied by a divergence of opinion about the evaluation of such counterfactuals, a divergence of opinion that can now be explained.

At first glance the effect of stochastic laws seems to be the same as that of ambiguity about which Cauchy surface to choose and vagueness in the instructions for changing the boundary data. In these cases the result is to generate a set of models instead of one, and the counterfactual only has a determinate truth value if the consequent obtains or fails to obtain in all of them. So too, if we precisely specify a Cauchy surface and the way data are to be changed, still stochastic laws will generate a set of models rather than one. Certainly, if the consequent obtains in all the models then the counterfactual is true, false if it obtains in none. And one can also maintain that the counterfactual has no determinate truth value if the consequent is true in some models and false in others. Michael Redhead, for example, takes just this view when discussing counterfactuals and quantum phenomena (Redhead 1987, pp. 90 ff.). As Redhead points out, this has a result which some consider counterintuitive. Suppose the world is governed by completely deterministic laws save for one that is stochastic: flipping a coin gives a 50 per cent probability of heads, 50 per cent tails. This probability is unaffected by all attendant circumstances. Further suppose that a flipped coin actually lands heads. What of the counterfactual: had I raised my arm a moment before the flip, it still would have come up heads?

On Redhead's analysis the counterfactual is not true, even though raising the hand may be acknowledged to have no causal influence on the coin. Following the recipe, we go back to a moment before the flip and change the data so that the hand is up rather than down. Letting the laws operate on the new data, we get two models, one with the coin heads the other with it tails. So application of the recipe yields results that have been endorsed by at least some philosophers.

The result is not entirely happy, though. It causes deep difficulties for any counterfactual analysis of causation. And it conflicts with what David Mermin has called the Strong Baseball Principle: 'The strong Baseball Principle insists that the outcome of any particular game doesn't depend on what I do with my television set—that whatever it is that happens *tonight* in Shea stadium will happen in exactly the same way, whether or not I am watching it on TV.' (Mermin 1990, p. 100). The Strong Baseball Principle is supposed to apply even if tonight's game involves some irreducibly stochastic events.

Applying the Strong Baseball Principle to the coin flip, whether or not I had watched the flip, it would have come out the same. Since raising the arm has by hypothesis no influence at all on the flip, we get the result that even had my arm been up, the coin still would have fallen heads.

Redhead's analysis, unless amended, conflicts not only with the Strong Baseball Principle but with the requirement that a subjunctive conditional whose antecedent turns out to be *true* reduce to a material conditional. Should my arm have been down, the coin would have come up heads—after all, my arm *was* down and the coin *did* come up heads. But if we follow the recipe, we don't get this result. We choose a Cauchy surface and find at Step 2 that the command has already been carried out and no changes need to be made. Allowing the laws to operate, we again get two models, one heads, one tails.

In deterministic cases, all occurrences are fixed by the boundary values and the laws. The tacit ceteris paribus condition applies only at Step 2: we are to effect the commanded change making as few collateral changes as possible. In stochastic cases, events are determined by boundary values, laws, and the random outcomes of stochastic processes. How are we to apply the ceteris paribus clause to this last element?

It is too strong to suggest that we should restrict our attention to models in which as many events as possible match those in the actual world. There is no determinate fact about how the coin would have fallen had we done something to it, such as flipping it higher or starting the flip in a different position. If there had been a different causal process leading to the result we might have gotten a different result. If the process is unchanged in the counterfactual, as in the case of raising my hand, so should the result be. The question is how we can determine which causal processes would be changed by a change in the boundary data.

A solution to this problem can only be sketched here, and perhaps never can be resolved by philosophical means. I hesitate to call what follows a theory; it is rather a description of how we may think about these cases. It explains why we might subscribe to the Strong Baseball Principle and it yields the result that if nothing had been changed, if a subjunctive conditional has a true antecedent, then the coin would be heads.

If I had raised my hand, the only results of stochastic processes that might have had different results are those that would have been (or might have been) affected by my hand going up. Any stochastic process is an evolving set of physical magnitudes. Let us call any physical magnitude which is unchanged when we apply Step 2 *uninfected*; those which are changed are, in both the original and new data sets, *infected*. This distinction may be quite clear in some theories, in others (notably in certain quantum states) the physical magnitude may not be localized and may be so entangled that the infection cannot be localized. In such cases, our intuitions break down.

The laws tell us how later magnitudes evolve from earlier ones. In some cases (again not in all, and perhaps not in quantum mechanics) a later magnitude can be seen as being generated from a set of earlier ones via the laws. Any magnitude generated at least in part from an infected magnitude is also infected. On one interpretation of the no-superluminal-signals constraint associated with Relativity, Relativity requires that all local magnitudes be generated from magnitudes exclusively in their back light-cones. In a pure particle theory with local interactions, infection could only be spread by contact with infected particles. So in at least some cases the generation of later states from earlier ones is of the right form to trace infection through time. In the actual world some of the later magnitudes will be infected, and in each of the models that results from the application of Step 3 the infection will spread in some way. Our ceteris paribus clause can now be stated: starting from the new Cauchy data and going forward in time, whenever the models that result from Step 3 start to diverge, if the divergence is due to the outcome of processes which, until that time, are identical and are uninfected in all of the models and in the actual world, then we should keep only those models in which the outcome agrees with the outcome in the actual world. If in any model the process has become infected, we must keep all the models generated by that process.

This prescription gives some good results. If the antecedent to a subjunctive is true, then Step 2 leaves all magnitudes uninfected, and we get back a single model that matches the actual world. If the magnitudes that are changed by watching rather than not watching TV do not propagate to Shea stadium, then the result of the game would have been the same no matter what I did. And if we deny this, saying that the magnitudes do propagate, then my intuition dissolves: perhaps I might have made a difference.

Although the infection theory sounds a bit complicated, it fits well one way of reasoning about counterfactuals. We start with the original situation and *update* it to fit the counterfactual condition. The first step in the update is Step 2 of the recipe. Then we consider what collateral changes those changes would have brought about. Our beliefs about the laws of temporal evolution guide this process. For an event that is remote from the region of the change, only a law connecting the events changed to that distant event would provoke us to update the model there. This is true even if the laws governing the situation are indeterministic. The difference between Redhead's and Mermin's intuitions is exactly the difference between those who *generate* a new model after Step 2 and those who *update* the original model on the basis of Step 2.[7]

The three-step recipe can also explain another sort of counterfactual: the counterlegal in which we assert what would happen were the laws to change. Step 2, rather than changing boundary values, alters instead the laws. Again a ceteris paribus condition accompanies the command: if told to change the gravitational constant we should leave the laws of electromagnetism alone. From the unchanged boundary conditions and the new laws we apply Step 3. The algorithm gives us the truth of 'If the gravitational force should disappear, the Earth would not continue to orbit the Sun.' All the problems of vagueness of the instructions reappear here.

Finally, we should note that although many counterfactuals involve consequents that occur later than the antecedent, some do not. If the bomb dropped on Hiroshima had contained titanium rather than uranium then a bomb containing titanium would have had to have been put on the plane. This can be accommodated in several ways. On the one hand, laws can in principle sometimes be used to retrodict as well as predict. In this case, one could use the recipe in unchanged form: change the Cauchy data and evolved backward rather than forward in time. But our intuitive sense of the world relies on LOTEs that typically work only in one direction. For example, given a glass of water with ice in it, on a hot day, we easily predict that if left alone the ice will melt and the water become lukewarm. But given a glass of lukewarm water on a hot day, and assured that it has been left alone over

[7] The distinction between updating and constructing does not make a difference if the laws are deterministic since only one model is consistent with the laws and the new boundary values.

the past hours, we would not hazard a guess about whether a few hours ago it contained ice. So whether or not the fundamental laws of physics are time reversible, the phenomenal laws we are familiar with are not. The natural tendency is not to evaluate counterfactuals whose antecedents postdate their consequents by 'running the laws backwards'—we wouldn't know how to do that—but rather by searching for the sorts of earlier states which would, in accord with the forward-running laws, lead to the later state. The only obvious way to *get* the titanium bomb onto the flying Enola Gay is to have had it *put* there earlier: one runs many scenarios forward in time rather than trying to run a single scenario (with the bomb on the plane) backward.

The vastness of the literature on counterfactuals does not allow an even comically inadequate comparison of this theory of (certain) counterfactuals with all the other proposals that have been made. I would, however, like to point out some contrasts with David Lewis's theory. Lewis's analysis postulates a context-dependent metric of overall similarity between worlds. Counterfactuals are evaluated by going to the nearest worlds in which the antecedent holds, and seeing if the consequent holds there. The theory presented here obviously has its roots in this approach. But Lewis's theory has been widely criticized on the grounds that no unprejudiced judgements of overall similarity yield the right result.

In deterministic contexts the problem is bad enough. Lewis concedes that in tracing out the consequences of a counterfactual condition in a deterministic world, the laws must be respected. In most details of particular fact, a world in which a titanium bomb is dropped *and explodes just like the actual uranium bomb* is much more similar to the actual world than one where it is dropped and doesn't explode. This is the wrong sense of similarity. So keeping the laws the same must take priority in judging the temporal evolution of worlds. On the other hand, what we call the changes due to Step 2 often require violations of law. We deflect the bullet without assigning a physical cause. Lewis wants these changes, so he admits that the nearest worlds will contain miracles, that worlds containing some miracles are closer to the actual world than ones that are thoroughly law governed. So what has become of the priority of law? Apparently, in the similarity metric some miracles are more equal than others.

Lewis employs some ingenious casuistry of miracle-comparison in an attempt to get the right result: some miracles are bigger or more widespread or more noticeable than others (Lewis 1986a, pp. 38 ff.). But none of those qualitative distinctions gets the right result: the miracles we don't care about are the ones *we* bring about in carrying out Step 2. *We* deflect the bullet because we are instructed to. Once we have the new data on the Cauchy surface (or on the pair of surfaces, one below and the other above the region which we change) then all that matters is the laws. And the changes that take place on that surface, or in the region between the pair, are justified by reference to the instructions, not by appeal to similarity. Indeed our recipe makes no reference to an overall similarity between worlds, the nearest thing being a ceteris paribus condition that determines what counts as the appropriate carrying out of a command.

If the problems in deterministic worlds are bad, those in stochastic worlds are insurmountable. It is not certainly true that had we started the coin flip in another position it still would have come up heads. But the model in which it does come up heads must be closer than the one in which it doesn't according to any similarity metric. For the two models match in point of lawfulness: the change in the boundary conditions is the same in both and each is an equally lawful, indeed equally probable, continuation of the boundary conditions. But the heads world is more similar to the actual world in point of particular fact, and if other events are decided deterministically on the basis of the flip, the differences in the two models may be massive. Lewis's analysis has no choice but to brand 'If we had started the coin in another position it would still have come up heads' true, but it clearly isn't. Our resolution of these cases by reference to infection makes no use of any comparisons of overall similarity.

Finally, we should note that Lewis's theory is concerned only with the semantics of counterfactuals, with the metaphysical conditions which make counterfactual claims true or false. He does not directly concern himself with the psychological question of how people evaluate counterfactuals, what processes underlie their intuitions. It seems quite unlikely that the psychological process could mirror Lewis's semantics: people do not imagine a huge multiplicity of worlds and subsequently judge their similarity and pick out the most similar. Rather we *construct* representations of possible worlds from the counterfactual description, construct them according to the three-step recipe or something like it. We imagine the proposed situation and let the scenario unfold in our mind, guided by the principles we accept as governing temporal evolution. Since the principal test of a semantic theory is how it accords with our intuitions, a semantics modeled on the process that generates intuitions is likely to be more satisfactory than one that ignores psychological procedure.

If counterfactuals are judged by a process resembling the three-step recipe then it is clear why, on the conceptual level, laws support counterfactuals.

We construct the counterfactual situation by means of the laws, so the laws must hold. On the ontological level, if the semantic values of counterfactuals are fixed by facts about laws in a way described by the recipe then the bearing of laws on the truth of counterfactuals is manifest.

6. MODELS, LAWS, AND EXPLANATIONS

There are two other views concerning laws and scientific theories that deserve our attention. They are in many ways opposed to one another, yet share similarities which permit them to be grouped together fruitfully for our purposes. The first is an account of laws championed by Storrs McCall and Peter Vallentyne.[8] The second is Bas van Fraassen's recently defended view that laws can be dispensed with altogether.

Vallentyne's theory begins with the notion of a world-history and an initial world-history. A world-history is 'a state of affairs that involves everything that happens at all points of time in some world'. An initial world-history is 'a state of affairs that involves everything that happens up to some point of time in some world, and involves nothing pertaining to later times. World-histories and initial world-histories involve only what happens; in particular, they involve nothing concerning nomic features of the worlds' (Vallentyne 1988, p. 604). A nomic structure is a relation between world-histories and initial world-histories. Roughly, it is the relation that holds between an initial world-history and a world-history iff the world-history is a continuation of the initial world-history that is permitted by the laws that govern the world. A law is any statement guaranteed to be true by the nomic structure.

Much could be said about the Vallentyne/McCall project, but the central issue is contained in the 'iff' sentence in the paragraph above. For Vallentyne it is an explication of the right-hand side by the left, whereas I have presented it as an explication of the left-hand side by the right. If we believe in FLOTEs (and adjunct principles) then we believe in models of FLOTEs (world-histories) and in a relation between models of a FLOTE and their initial segments. On this account, laws are primitive and nomic structures can be defined from and explained in terms of them. On Vallentyne's account it is the other way round.

[8] As with the Mill–Ramsey–Lewis theory and the Armstrong–Dretske–Tooley theory, I will use the writings of one representative, in this case Vallentyne 1988.

I certainly cannot criticize any account for having primitives that are not further explicated. But I do think one can pass judgement on the intuitive clarity of the primitives. Anyone who has studied physics or biology or chemistry or economics has a notion of what a law is. Anyone who has watched plants grow understands what it is to take 'plants grow toward the light source' as a non-accidental fact about plants, a feature of their development that allows us to predict what a plant will do and to judge what it would have done. It is from individual facts like these that we build up a picture of how the world might go, what the world-histories might be, not vice versa. It is very hard to see what a *primitive* grasp of a nomic structure might be, a grasp that might be obtained before the notion of law is in play.

It is also highly dubious that sense can be made of a world-history antecedent to the notion of a law. Laws and the physical magnitudes they govern are to some extent mutually determining. It makes little sense to suppose that an electromagnetic field exists in a region of space if the object in that region does not at least approximately obey Maxwell's laws, does not deflect charged particles, etc.[9] So if world-histories involve claims about electromagnetic fields and charged particles, then the laws must already be settled to a large extent by the world-history before any question about a further relation between the world-history and its initial segments can be raised. When Vallentyne says that world-histories 'involve nothing concerning the nomic features of the worlds' it is unclear what sorts of states of affairs, if any, can meet this requirement.

Finally there is the question of strength. We have already seen that an account that takes FLOTEs as primitive can do all of the work of Vallentyne's view since the nomic structure can be defined from the FLOTEs. Can we inversely define the laws, and particularly the FLOTEs, given only the nomic structure?

Laws are supposed to be propositions whose truth is guaranteed by the nomic structure, propositions true in all the world-histories with a given nomic structure. The FLOTEs should certainly turn out to be laws (I don't know how to prove this except by generating the world-histories from the FLOTEs because I don't know how else to identify the relation in question). But are all the other propositions whose truth is guaranteed by the nomic structure laws?

They are not if one of the classic criteria for laws is accepted: confirmation by positive instances. By Vallentyne's criterion, 'Everything which is either a

[9] In tribute to Wilfrid Sellars's parallel claim about concepts, we might dub this doctrine 'Physical Magnitudes as Involving Laws and Metaphysically Impossible without Them'.

cow or a sample of Uranium 238 is either a mammal or radioactive' is a law. But observations of any number of one sort of positive instance (cows) gives us no reason to believe anything about others (uranium). Confirmation is not transmitted to all unobserved instances. It is natural to regard the claim above as a consequence of two laws which is not itself a law. Vallentyne's definition does not allow this.

More controversially, I would object to calling certain propositions laws even though they are confirmed by their positive instances. 'All humans who live in houses with prime house numbers are mortal' is not a law because the class referred to is not a natural kind. Positive instances do transmit confirmation to unobserved cases, but only because they confirm 'All humans are mortal,' and all instances of the first claim are instances of the second. The proposition is not a law but again a consequence of a law, this time by conjunction rather than disjunction.

Even more controversially, I don't think that 'All humans are mortal' is a law of nature even though it is couched in terms of natural kinds and properties, confirmable by positive instances, and guaranteed by the nomic structure. It too is a consequence of laws, laws of biochemistry, physiology, etc., but is not itself a law.

Being purely cranky, one could challenge the advocate of the lawhood of 'All humans are mortal' to find the generalization treated as a law, or even stated, in a biology text. One might find it remarked that humans are mammals, omnivores, etc., but not that they are uniformly destined to die. The equally cranky response would be that of course one never writes that all humans are mortal because it *goes without saying* that all humans are mortal. Everyone knows it. So why waste ink printing it? But the supposed law is not missing just because it is tacit. No biological fact is ever *explained* by reference to this law, whether the reference be tacit or explicit. In particular, no one's *death* is ever explained by reference to this law. People die of cancer, or stroke, or trauma, or asphyxiation. Nobody ever dies of humanity.

At the beginning of this article I remarked that one of the widely acknowledged conceptual connections is between laws and explanations. The covering law model of explanation took this connection as its primary inspiration. We need not review the shortcomings of this model: not every subsumption under law, not every derivation of a fact from laws and initial conditions, constitutes an explanation. Nonetheless, there are distinctive explanations whose primary structure consists in showing how the operation of laws of nature led to a particular event or state of affairs. We understand why the planets obey Kepler's laws (as well as they do) when we derive the trajectories of bodies with certain initial velocities and masses from the General Relativistic laws. The operation of a FLOTE explains why certain physical magnitudes take on values at later times given their values at earlier times. Not every subsumption is an explanation, but a general proposition that cannot be used to explain any of the instances it covers can hardly be called a law.

Of course, given the right context the statement 'All humans are mortal' could play a role in providing understanding of a fact. But given the right circumstances so could 'All the coins in N.G.'s pocket on VE Day were silver.' The issue is whether the explanatory power derives merely from the truth of the assertion or from its status as a law. The explanation of individual cases by subsumption does depend on lawhood, for it is acknowledged that subsumption by accidental generalizations does not explain. In the 'all heads' world governed by stochastic laws it in no way explains why a particular coin came up heads to note that they all did. Indeed, since the laws are by assumption irreducibly stochastic, there *is* no explanation of why a particular coin came up heads. In the deterministic 'all heads' world, the particular events are explained by the law. We might, of course, seek a further explanation for the law.

This picture of explanation by nomic subsumption faces a severe challenge in Bas van Fraassen's *Laws and Symmetry* (1989). Van Fraassen urges us to abandon all philosophical accounts of laws as unworkable, and to abandon laws altogether, seeking an account of scientific practice and of theories that eschews all talk of laws. Much of van Fraassen's inspiration derives from the observation that many laws (especially conservation laws) that had been postulated as primitive in classical physics are now derived from underlying symmetries. This aspect of van Fraassen's view will not come in for further notice here.[10] What are of interest are his views on scientific theories. Since theories are often taken to be attempts to formulate the laws of nature, van Fraassen must come up with an alternative account.

His solution lies in a particular interpretation of the semantic conception of scientific theories. Although theories are often presented as (more or less) axiomatized systems from whose axioms the models can be determined, van Fraassen sees the axioms or formulation of laws as merely accidental artefacts.

[10] Except for this: The connection between symmetries and conservation laws is a consequence of Noether's theorem, which shows that under certain conditions every symmetry of a Lagrangian implies a conserved current. But Noether's results are only relevant for systems governed by a Lagrangian, i.e. systems that obey the Euler–Lagrange equations, and this is a matter of being governed by certain sorts of laws.

The content of the theory is not in the laws but in the set of models one ends up with: 'if the theory as such is to be identified with anything at all—if theories are to be reified—then a theory should be identified with its class of models' (van Fraassen 1989, p. 222).

Van Fraassen adds to this view his own peculiar account of the objectives of theorizing. Famously, he believes that science is not a search for theories which are true, i.e. which contain a model which correctly describes the entire actual world, but rather theories that are empirically adequate, i.e. which contain a model which fits all of the observable phenomena in the actual world.

Now science cannot aim *merely* at theories that are empirically adequate, for that is trivially accomplished. The 'anything goes' theory has as models all logically possible distributions of observable phenomena through space-time, and so can accommodate itself to anything we see. Such a 'theory' would not explain anything, and could not be used to predict anything, but it would be empirically adequate.

For the same reason, science cannot seek merely theories that are *ontologically adequate*, i.e. that contain a model that is isomorphic to the whole actual world, observable and unobservable. We seek theories that are small classes of models; at least we want to cut down the models of the 'anything goes' theory. The questions are how this is done and why this is done. What are the principles that govern the acceptance of ever stronger theories?

One might maintain that we seek theories that are *metaphysically adequate*, i.e. theories whose models stand in one-to-one correspondence with the physically possible states of affairs, each model being isomorphic to a state. On the view that we seek theories that correctly state the laws of nature, this is true. But it is unavailable to van Fraassen unless he reifies the physically possible worlds, and is unavailable on general empiricist grounds even then. So we are still left with the questions: how do we slim down our theories and how does their explanatory power increase in virtue of reducing the class of models?

If a theory is just a class of models then the only solution that appears (to me) to be possible is that a theory explains a fact in the actual world only if that fact obtains in all (or an appropriately defined most) of its models. We can state this position without reference to laws and hence without reference to explanation by nomic subsumption. But it is easy to see that if we let the models play the role of Vallentyne's world-histories then we can adopt his definition of law: a law is a proposition that is true in all the models. Then the proposed account of explanation by theories amounts to explanation by subsumption under Vallentyne-laws. Van Fraassen would have to regard this way of putting the matter as innocuous since we have defined Vallentyne-laws from his primitives. We now have a rationale for improving on the 'anything goes' theory: theories with fewer models can explain more.[11] But this account buys explanation too cheaply. For we must take quite seriously the idea that the theory *is just* the class of models. The mode of presentation, the 'derivation' of the models from laws and boundary conditions and choices of results for stochastic processes, is inessential. Consider the theory that results if we begin with our present theories of physics, chemistry, etc. but exclude all the models which do not contain living beings. This is a perfectly good class of models, and hence a theory. In this theory, the existence of life in the universe is a 'law'. Hence the existence of life in the universe is explained as the consequence of a 'law'. If our old theory was empirically adequate or ontologically adequate, so is the new one. Hence the new theory is preferable to the old one: it is as likely to be adequate as the old and more explanatory. And clearly the same trick can be played for any known fact: simply restrict the class of models from the old theory to those in which the fact holds, and thereby produce a theory that explains it. This does not describe actual practice.

Since science would not rest content with the 'anything goes' theory, the drive for, and constraints on, the search for logically stronger theories must be explained. It is hard to see how to explain this drive if one takes the class of models as primitive and the means of presenting or defining the class as incidental. On the view defended in this essay models are derivative, laws primary. The content of a particular model is determined by three factors: the laws, the boundary values, and the results of stochastic processes. Correspondingly, there are three kinds of regularity in a model: those due entirely to the laws, those due in part to the boundary conditions, and those due in part to the results of stochastic processes (the last two classes may overlap). Regularities of the first kind are explicable by nomic subsumption, the other kinds are accidental and have no explanation.[12] There is no explanation in the stochastic 'all heads' world for the regularity of results of coin flips. It is just by chance.

[11] I have here ignored much of van Fraassen's account of the pragmatics of explanation in his (1980), but they are not relevant for the point. Once the various context-dependent features of the explanatory context have been determined, one still must explain why logically stronger theories can explain more than logically weaker ones.

[12] This formulation is too strong, but is a good starting point. As we will see below, there are degrees to which a regularity can depend on particular initial conditions or on fortuitous results of random processes. If the degree is sufficiently reduced we consider the regularity explained.

So far we have dealt only in toy cases. But scientific practice supplies striking examples of the search for explanation of regularities by nomic subsumption and the role that search plays in evaluation of theories. The next section will put some meat on these logical bones by examining a contemporary example.

7. *EXEMPLI GRATIA*: THE INFLATIONARY COSMOLOGY[13]

If one looks into the sky in the microwave range one sees, in all directions, an extremely uniform background radiation with an effective temperature of about 3° Kelvin. The temperature of the radiation deviates no more than one part in 10,000 in different regions of the sky. If one calculates the ratio of the energy density of the universe to the critical density, i.e. the density which divides opens universes (which continue expanding forever) from closed universes (which end in a Big Crunch), one finds that it is near 1: current calculations yield a range of values from 0.1 to 2. If one looks for magnetic monopoles, massive particles with an isolated magnetic pole, one does not readily find one; indeed no uncontroversial sighting has yet been reported. If there are any, there are few.

Standard Big Bang cosmology is ontologically (and hence empirically) adequate with respect to these facts. There are models of Big Bangs which have as uniform background radiation as one likes, that have energy densities arbitrarily close to critical density, that have few or no magnetic monopoles. Nonetheless, these phenomena are considered to pose deep, perhaps fatal, objections to standard Big Bang cosmology. Why? To avoid the suspicion that a philosophical interpretation is being imposed on this case study, I quote at length the cosmologists Alan Guth and Paul Steinhardt:

When the standard big-bang model is extended to these earlier times [i.e. between 10^{-43} and 10^{-30} seconds after the singularity], various problems arise. First, it becomes clear that the model requires a number of very stringent assumptions about the initial conditions of the universe. Since the model does not explain why this set of initial conditions came to exist, one is led to believe that the model is incomplete—a theory which included a dynamical explanation for these conditions would certainly be more convincing. In addition, most of the new theories of elementary particles imply that exotic particles called magnetic monopoles (each of which corresponds to an isolated north or south magnetic pole) would be produced in the early universe.

[13] This presentation follows that of Guth and Steinhardt 1989, which is highly recommended to those interested in a non-technical presentation of the physics.

In the standard big-bang model, the number of monopoles would be so great that their mass would dramatically alter the evolution of the universe, with results that are clearly inconsistent with observations.

The inflationary universe model was developed to overcome these problems. The dynamics that govern the period of inflation have a very attractive feature: from almost any set of initial conditions the universe evolves to precisely the situation that had to be postulated as the initial state in the standard model. Moreover, the predicted density of magnetic monopoles becomes small enough to be consistent with the fact that they have not been observed. In the context of the recent developments in elementary particle theory, the inflationary model seems to be a simple and natural solution to many of the problems of the standard big-bang picture. (Guth and Steinhardt 1989, p. 34)

Why does the standard theory have such problems? First microwaves. According to the standard scenario the universe expanded from the initial singularity at a relatively constant rate (on the appropriate logarithmic scale). If we retrodict using this model of expansion we find that the regions from which the microwaves originated were never in causal communication with each other (if no cause operates at greater than the speed of light). These regions could therefore never have come into thermal equilibrium. The radiation that comes to us from the north traces back to a different part of the initial boundary conditions than that which comes from the south, and the processes by which the radiation has propagated have never been in causal contact. So the only way for the radiation to have precisely the same temperature in both places is for the initial boundary conditions to be already highly uniform. Of all the distributions of values that the independent variables in the initial state could take, they just happened to take an almost perfectly uniform distribution. Thermal equilibrium could account for uniformity over small patches of the sky, but uniformity over the whole sky must be put in by hand, as a very special choice of initial conditions.

Similarly, the energy density of the universe must be put in by hand. In the standard model this parameter is a free variable that could take any value. Furthermore, in the standard dynamics the critical density is a point of unstable equilibrium.

If W [the ratio of the energy density of the universe to the critical energy density] was ever exactly equal to one, it would remain exactly equal to one forever. But if W differed slightly from one an instant after the big bang, then the deviation would grow rapidly with time. Given this instability, it is surprising that W is measured today as being between 0.1 and 2. (Cosmologists are still not sure whether the universe is open, closed, or flat.) In order for W to be in this rather narrow range today, its value after the big bang had to equal one within one part in 10^{15}. The standard model offers no explanation of why W began so close to one, but merely assumes the fact as an initial condition. (ibid. 37)

Finally, magnetic monopoles. According to present theory, the universe contains a field, the Higgs field, with the following peculiar property. When the field is uniformly zero (the 'false vacuum'), it is not in its lowest energy state. The false vacuum can decay down into a lower energy state (the 'true vacuum'), liberating energy in the process. In the true vacuum, the Higgs field takes on a non-zero value. Indeed, there are *many* true vacuum states, in each of which the Higgs field has a *different* value. The process of decay is called *spontaneous symmetry breaking* since the Higgs field goes from a symmetrical false vacuum state to a particular true vacuum that does not display symmetry under transformations of the field. The process of spontaneous symmetry breaking is random, being governed by quantum laws. We cannot predict when it will occur or, more importantly, into which true vacuum state the field will decay.

After the Big Bang, the Higgs field will be in a false vacuum state. As the universe cools, it becomes energetically favorable for spontaneous symmetry breaking to occur. Once the field decays at a certain point, that value of true vacuum can spread out from the point, but again not faster than light speed. So the same problem as appeared with the radiation appears again. There can be no mechanism to ensure that regions that are not in causal communication end up in the same true Higgs vacuum state. Regions outside the light-cone 'horizon' of one decay event will decay independently, and it is almost unimaginably unlikely that they will arrive at the same true vacuum state.

What happens when two regions with different true vacuums collide? The incompatibility of the vacuum states causes defects to appear where the Higgs field is not well-defined. Linelike defects are called *vortices* or *cosmic strings*; planelike defects are *domain walls*. Pointlike defects are *monopoles*. The more independent decay events there are, the more defects of all kinds there should be. According to the standard cosmology, many decay events should have occurred because the visible universe evolved from many causally unconnected regions. But few if any monopoles exist.

To repeat, it is not that the standard cosmology cannot accommodate the observed facts. We can choose initial conditions that yield smooth and uniform background radiation; we can set W as near as we like to one; we can even assert that the various random decay events all happened to decay to the same value, or that there has only been one decay event, so there are no defects. But these occurrences are monstrously unlikely. Most initial conditions give a universe far from the critical energy density with widely varying background radiation.[14] The quantum mechanical probability of having unnoticeably few monopoles is incalculably small. Despite all that, the standard model is ontologically adequate.

But as Guth and Steinhardt insist, although the standard cosmology can *accommodate* these facts, it cannot *explain* them. The observed regularities would be due to primitive regularities in the initial data or to remarkable fortuitous results of random processes, and this does not constitute an explanation. We want a *dynamical* explanation, a theory that traces these regularities back to the operation of laws. The phenomena should appear independently (or nearly independently) of initial conditions and should not need conspiracies among random processes.

Roughly, here is how the inflationary theory accomplishes this. Recall that we don't have any trouble explaining why a region of the visible universe that traces back to a small enough segment in the early universe has uniform radiation or few monopoles. In a small enough segment, the different places are in causal contact with one another. So the radiation can come into thermal equilibrium and the Higgs vacuum can spread to fill the whole space. The problem is that in the standard picture the visible universe expanded from many causally unconnected regions, not from one connected one.

On the inflationary model, the universe at 10^{-43} seconds was at least some fifty orders of magnitude smaller than on the standard picture, so small that the whole of the visible universe could trace back to one causally connected region. At about 10^{-35} seconds a period of hyper-exponential inflation set in, driven by the decay from the false vacuum to the true vacuum. Between 10^{-35} and 10^{-33} seconds the universe grew by a factor of 10^{50}. After that the Higgs field freezes out in the true vacuum and standard Big Bang cosmology takes over.

As a result, the microwave and monopole problems are solved. The distant ancestors of the present microwave fields were in thermal equilibrium, so they had to come to the same temperature. Only very few events of spontaneous symmetry breaking occurred. And as a bonus, the hyperinflation tends to 'smooth out' the curvature of the universe, driving it towards the critical value of energy density rather than away. From almost any initial conditions

[14] This 'most' requires some measure over the space of initial conditions. One can squabble about the exact form of a measure, but for any reasonable, intuitive measure the space of conditions leading to the observed regularities is minuscule.

the dynamical laws deliver a universe nearly flat, with constant radiation, and with few defects.

It is clear in what sense the inflationary model explains the features that the standard model does not. The phenomena do not follow from the laws alone, but in the inflationary model *avoiding* the observed results requires the same fine-tuning of initial conditions, the same fortuitous results of random process, that the standard model needs to *accommodate* the phenomena. The inflationary laws render the results massively insensitive to changes in initial data and give them a high objective probability. It is fair to call this explanation by subsumption under law.

We can understand this judgement of explanatory power if we recognize the role of laws in generating models. The data are explained by the laws because the data obtain in most of the models generated by the laws together with the variety of possible initial conditions and results of random processes. In most of the physically possible universes that develop according to these laws, the radiation is uniform, etc. The effects are explained by the laws because they occur in most of the models generated by the laws.

But we cannot understand these judgements of explanatory power on the semantic view as van Fraassen espouses it. If we care only about the set of models of the theory and not about the laws that generate them, then the only way in which the inflationary theory can be said to explain the data is that the data occur in most of the models. But if that is all that was needed, an explanatory theory was always near to hand. Simply take the models of the standard cosmology and throw out all the models with non-uniform radiation, with energy densities far from the critical value, or with many magnetic monopoles. In this new theory, this new class of models, the troubling phenomena occur in most of the models. But this ridiculous ad hoc procedure would not provide a theory which explains the phenomena. Indeed, if one accepts a semantic view according to which formulations of laws are only inessential means of specifying classes of models, no sense at all could be made of the demand of Guth and Steinhardt for a *dynamical* explanation. The dynamics depend on the laws, so an account of scientific practice that can accord with this episode must allow laws a non-incidental role in theories.

8. PERORATION

In many respects this proposal about laws of nature is modest in just the way Swift's was. I have urged a radical reconstruction of the problematic surrounding laws, shifting them from analysandum to analysans. With laws in hand, connections to possibilities, to counterfactuals, and to explanations come swiftly and easily, with better results than many other philosophical approaches yield. But in other ways, this proposal is truly modest, a sketch to be elaborated. I would like to indicate some directions I think this elaboration should take, and some respects in which modesty will ever remain a virtue.

This chapter has concentrated on *fundamental* laws of temporal evolution, with examples drawn from physics. The mark that we believe a law to be a fundamental law is that exceptions are not tolerated. If the law describes the pattern that nature respects in generating the future from the past then those patterns must appear wherever the law is in force.

The laws of the special sciences do not aspire to this status. Biology, chemistry, psychology, geology, as well as our everyday rules of thumb about how things go, all involve descriptions of how systems typically evolve, but exceptions of certain sorts are both tolerated and expected. Not every embryo follows the usual developmental pattern, not every gene pool evolves according to the laws of population genetics. To take the most blatant example, the laws of population genetics will not govern the evolution of a species subject to nuclear holocaust. An acid mixed with a base will not form a salt if the student knocks the beaker off the table.

Laws that admit of exceptions are sometimes called *ceteris paribus laws*, and are seen as posing philosophical difficulties. The term 'ceteris paribus' is inapplicable here, though, absent any reference class in comparison with which we can determine whether the *cetera* are *paria*.[15] Since these are typically laws of the special sciences, let us call them instead special laws. Some sciences, such as population genetics, contain both Special Laws of Temporal Evolution (SLOTEs) and special adjunct principles.

What is the logical form of a special law? How do special laws connect with possibility, counterfactuals, and explanation?[16]

My proposal is this: the logical form of a special law is just like the logical form of a fundamental law with the addition of the clause 'if nothing interferes'. Population genetics, if true, tells us how gene pools will (or might) evolve if nothing interferes. The connections to possibility, counterfactuals,

[15] We may contrast this with the *ceteris paribus* condition that applies at Step 2 of the recipe: there the actual world provides the standard of comparison. Of course, we must still decide which *cetera* are to be kept *paria*.

[16] The following thoughts on special laws have been greatly influenced and stimulated by discussion with Brian McLaughlin.

and explanations are taken over *in toto* to the special sciences. The laws of population genetics beget models that describe how gene pools evolve if nothing interferes. They support counterfactuals about what would or might happen in various circumstances if nothing interferes. They explain features of genetic evolution when nothing interferes. And if the laws of population genetics are true, then actual populations will evolve in accord with them if nothing interferes.

Now all of our suspicions have become focused on the rider 'if nothing interferes'. One line of thought maintains that as it stands the rider is empty; it can only be given content if all of the possible forms of interference are explicitly enumerated. But now the laws of chemistry or biology must become bloated with admonitions about nuclear explosions, meteor collisions, clumsy students, etc. And no matter how bloated it gets, some possible interfering factors will always be left out.[17]

One further suspects that 'if nothing interferes' can be used as a weasel clause to save bad theories. Telepathy is alleged to be a highly fragile phenomenon with everything from sunspot to skeptics producing bad vibes that destroy the effect. Isn't such an escape clause the mark of pseudo-scientific wimps who won't face the tribunal of experience?

And lastly, one might maintain that such a clause robs a theory of its explanatory power. Traditional theories of explanation discerned a parallelism between explanation and prediction. But if the content of the interference clause is not made explicit we can no longer predict anything with confidence. We won't know if we have controlled for all possible forms of interference. The best we can do is to make a conditional prediction: if nothing interferes, then …

This last objection surely depends on a criterion of explanation that is too strong, for it would rule out most scientific practice. No chemist would predict the result of an experiment if we admonish her to take the likelihood of nuclear war into account. No psychologist would be surprised if a subject failed to act as predicted if in the interim the subject had been struck by lightning; nor would the psychologist feel obliged to take up meteorology. Telepathic phenomena may claim to be fragile, but so is scientific instrumentation. If the experimental reports of one lab are not verified by a second lab following the

[17] In some cases, all interference may be eliminated by restricting the law to hold for *isolated systems*, where an isolated system is one surrounded by a surface through which no energy or matter is allowed to pass. Unfortunately, no actual systems are isolated in this way.

most exact instructions we can concoct, we do not automatically reject our theory. There might be a bug in their apparatus, a bug of a sort that is not familiar.

No scientist will predict with certainty that a new instrument will work as it was designed to. But if it does so work, its behavior can be explained by the laws used to design it. The original prediction tacitly contains the rider: such-and-such will happen (if nothing interferes). If such-and-such *does* happen, and is not antecedently likely (e.g. if a sharp image of a binary star system appears), then one infers (fallibly) that nothing did interfere. And if nothing interfered, the whole battery of special laws shed their interference clauses and can be used for explanation as if they were fundamental.

The important case is that in which the instrument fails to behave as predicted. Since the laws imply that it will so behave if nothing interferes, if it doesn't work (and we still believe the laws) we infer that something interfered. The pseudo-scientist stops here, taking credit for the successes and fobbing off the failures. But this attitude is not warranted by the law. The conclusion that something interfered is a contingent claim, and as such we must seek to verify it. A bug in the apparatus ought to be able to be found and corrected. If sunspots caused the problem then shielding should eliminate it. If there was interference then there must be some event or object such that had it been removed the expected result would have ensued. This last counterfactual is, of course, underwritten by the FLOTEs.

We have already seen that some pragmatic considerations beyond the laws themselves enter into the evaluation of counterfactuals. To this extent, SLOTEs may be less objective than FLOTEs, for the SLOTE is true only if counterfactuals about the removal of interfering circumstances are true. Furthermore, there may be perfect agreement about under what circumstances the SLOTE would obtain (i.e. give the predicted result) but still disagreement about which of those circumstances constitute the presence or absence of interfering factors.

In many cases, such as that of the student knocking the beaker off the table, it is easy to locate uncontroversial events that constitute interference. And there are equally clear cases where interference does not exist. Suppose one were to attempt to defend as a psychological law the claim that all people, when thirsty, stand on their heads (if nothing interferes). Or worse, that they float to the ceiling (if nothing interferes). These laws are confronted by an uncounted multitude of prima facie counter-examples. It would be safe to say that there are no factors that would be considered interfering factors such that, had they been removed, the predicted consequent would have ensued.

So special laws are not reduced to being empty by the interference clause, even though in some cases the application of the clause may be moot. If this approach is anywhere near being on the right track, the conditions under which we judge that interference has occurred are worthy of further study.[18]

The deepest sense in which this essay has been modest is the degree to which central notions, such as 'law of temporal evolution', 'transmission of infection', and now 'interference', have been left intuitive and vague. There are clear examples of the application of these concepts, and in exactly the cases where these concepts have clear application our intuitions about possibility, counterfactuals, and explanations are clearest. This suggests we are on the right track. But I have provided no rules for the extension of these notions to new contexts.

To some extent we are in a position to do this. For example, our paradigms of FLOTEs, Schrödinger's equation, and Newton's laws are defined for non-relativistic contexts. The idea of a law governing temporal evolution can be extended to special relativistic, and (I think) even general relativistic, contexts. But no philosopher could have predicted these theories before their formulation and allowed for them. Future theories may challenge the notion of time itself, or of a physical state. If space-time becomes a quantum foam at small length scales, the idea of temporal evolution may break down there. But it is the role of philosophers to study and learn from the radical reconceptions of physical reality, not to dictate to them. This sort of vagueness about notions such as 'law of temporal evolution' is, if not a virtue, at least a necessary evil.

Some possible laws would occasion revisions to the explications offered in this paper. For example, a temporal-action-at-a-distance law, according to which some events influence events at later times without having any effect on the intervening physical states, would entail a reworking of the idea of a Cauchy surface, and a corresponding change in the three-step recipe. More radical suggestions may yield situations where the recipe is inapplicable even with revision, and must be given up.

But the mark of a successful philosophical account of the sort offered here is not the universal applicability of its primitive concepts under every circumstance. The mark is rather that when the primitive concepts fail or become fuzzy, so should intuitions about all the other notions that have been built on them. One last example may illustrate this point.

[18] Such a study would doubtless begin by delineating *degrees* of interference (and hence degrees of isolation). Probably any actual biological process is affected to some degree by interfering events (e.g. the passage of cosmic rays). Usually such events only slightly perturb the system and so can be ignored if one is interested in gross phenomena.

One analysis of counterfactuals in stochastic contexts relied on the idea of infection. We supposed that the physical magnitudes that are changed at Step 2 can be identified and the subsequent physical magnitudes that depend on them distinguished. This picture is clearly applicable when the physical state consists in locally defined physical magnitudes and the laws are laws of local interaction. In that case, the infection spreads out from an initial locus and can be traced.

In quantum theory both of these conditions may fail. The famous 'entangled states' of a correlated pair of particles cannot be understood as the sum of local states of each particle. The laws describe the evolution of a vector in Hilbert space, not the evolution of spatially local magnitudes. And on the Feynman path integral picture, every particle travels along every possible trajectory, so infection would immediately spread everywhere.

And in exactly the cases of correlated quantum pairs, intuitions about counterfactuals break down. The disputes over the so-called 'Principle of Counterfactual Definiteness', used by Stapp and Eberhard in their proof of Bell's theorem, illustrate this (see Redhead 1987, pp. 90 ff.). The weirdness of quantum theory arises in part from its introduction of concepts that defy easy analysis in terms of the notions that underpin our judgements of counterfactuals, locality, and causality.

The notions of law, possibility, counterfactual, and explanation are deeply interconnected. The directions of the connections ought to be open to all hypotheses. If we take laws as primitive, relatively clean analyses can be given of the rest, analyses that fit intuitions, predict degrees and sorts of context dependence, and describe actual scientific practice, at least in some notable instances. Shortcomings of other theories are avoided and their failures explained. Of course, a polemical essay such as this contains only the happiest examples and competing theories are not given the benefit of every doubt. But I hope that at least a first foothold has been carved for the idea that laws are primitive, primitive components of our conceptual system and even primitive elements of physical reality.