former worlds, but the assertability of counterfactuals depends on the character of the latter worlds.*

Perhaps I have considered the wrong thought experiment; the right one is to add the antecedent to your system of beliefs not as if it were an item of new knowledge, but simply *as a counterfactual supposition.* That is the right thing to do, I agree, but it is unhelpful to say so. For what is the thought experiment of adding ϕ to your beliefs as a counterfactual supposition? I suppose it is nothing else than the exercise of deciding which counterfactuals with the antecedent ϕ you believe.

3.3 The Metalinguistic Theory: Laws of Nature

Metalinguistic theorists commonly give a special place among cotenable factual premises to laws of nature. A law is thought to be cotenable with any antecedent, except an antecedent that is logically inconsistent with that law, or perhaps with some other law. On this view, if the antecedent of a counterfactual, together with some laws, implies the consequent, and if the antecedent is logically consistent with all laws, then the counterfactual is true. (Or: if that is thought to be the case, then the counterfactual is assertable.) On this view also, there can be no true counterfactual saying that if so-and-so particular state of affairs were to hold, then such-and-such law would be violated.

I could, if I wished, incorporate this special status of laws into my theory by imposing the following constraint on systems of spheres: the set of all and only those worlds that do not violate the laws prevailing at

* Perhaps the method of thought experiments gives the proper conditions of assertability not for counterfactuals but for *indicative* conditionals. Let P be a subjective probability function representing your system of beliefs, and let P_ϕ likewise represent the revised belief system that would result under the new item of knowledge ϕ. According to standard Bayesian confirmation theory, $P_\phi(\psi) = P(\psi / \phi) =^{\text{df}} P(\psi \,\&\, \phi) / P(\phi)$, provided that the denominator $P(\phi)$ is nonzero. Ernest Adams has observed that $P(\psi / \phi)$ seems also to measure the assertability of the indicative conditional '*If* ϕ, *then* ψ' according to the belief system P; see his 'The Logic of Conditionals', *Inquiry* **8** (1965): 166–197. For instance, I do assert that if Oswald did not kill Kennedy, then someone else did; I do so because P (*Someone else did / Oswald did not*) is high. And I deny that if Oswald did not kill Kennedy, then Kennedy was not killed; that is because P(*Kennedy was not killed / Oswald did not kill him*) is low. (The indicatives are thus opposite in assertability to the corresponding counterfactuals.) Adams's observation about the *assertability* conditions for indicative conditionals is compatible with various alternative views about their *truth* conditions, or lack thereof. I favor the view that the indicative conditional '*If* ϕ, *then* ψ' has the truth conditions of the material conditional $\phi \supset \psi$; its assertability is measured by $P(\psi / \phi)$ rather than $P(\phi \supset \psi)$ because if the latter is high and the former is low, then $P(\sim\phi)$ is almost as high as $P(\phi \supset \psi)$ and it is pointless and misleading to assert $\phi \supset \psi$ rather than $\sim\phi$.

a world *i* is one of the spheres around *i*. Equivalently, in terms of comparative similarity: whenever the laws prevailing at *i* are violated at a world *k* but not at a world *j*, *j* is closer than *k* to *i*. This would mean that any violation of the laws of *i*, however slight, would outweigh any amount of difference from *i* in respect of particular states of affairs.

I have not chosen to impose any such constraint. I doubt that laws of nature have as much of a special status as has been thought. Such special status as they do have, they need not have by fiat. I think I can explain, within the theory already given, why laws tend to be cotenable, unless inconsistent, with counterfactual suppositions.

I adopt as a working hypothesis a theory of lawhood held by F. P. Ramsey in 1928: that laws are 'consequences of those propositions which we should take as axioms if we knew everything and organized it as simply as possible in a deductive system'.* We need not state Ramsey's theory as a counterfactual about omniscience. Whatever we may or may not ever come to know, there exist (as abstract objects) innumerable true deductive systems: deductively closed, axiomatizable sets of true sentences. Of these true deductive systems, some can be axiomatized more *simply* than others. Also, some of them have more *strength*, or *information content*, than others. The virtues of simplicity and strength tend to conflict. Simplicity without strength can be had from pure logic, strength without simplicity from (the deductive closure of) an almanac. Some deductive systems, of course, are neither simple nor strong. What we value in a deductive system is a properly balanced combination of simplicity and strength—as much of both as truth and our way of balancing will permit. We can restate Ramsey's 1928 theory of lawhood as follows: a contingent generalization is a *law of nature* if and only if it appears as a theorem (or axiom) in each of the true deductive systems that achieves a best combination of simplicity and strength.‡ A generalization is a law at a world *i*, likewise, if and only if it appears as a theorem in each of the best deductive systems true at *i*.

In science we have standards—vague ones, to be sure—for assessing

* See 'Universals of Law and Fact', in Ramsey, *Foundations*. (R. B. Braithwaite kindly permitted me to see this note in manuscript.) Ramsey regarded it as superseded by 'General Propositions and Causality', also in *Foundations*. He there alludes to his previous theory of 1928 in the words I have quoted (page 138); rejects it on the ground that we never will know everything; and goes on to develop a different theory. See also Braithwaite's mention of the 1928 note in his editorial introduction, *The Foundations of Mathematics*: xiii.

‡ I doubt that our standards of simplicity would permit an infinite ascent of better and better systems; but if they do, we should say that a law must appear as a theorem in all sufficiently good true systems.

the combinations of strength and simplicity offered by deductive systems. We trade off these virtues against each other and against probability of truth on the available evidence. If we knew everything, probability of truth would no longer be a consideration. The false systems would drop out, leaving the true ones to compete in simplicity-cum-strength. (Imagine that God has decided to provide mankind with a *Concise Encyclopedia of Unified Science*, chosen according to His standards of truthfulness and our standards of simplicity and strength.) Our standards of simplicity and strength, and of the proper balance between them, apply—though we who are not omniscient have no occasion so to apply them—to the set of all true deductive systems. Thus it makes sense to speak of the best true systems, and of the theorems common to all the best true systems.

I adopt Ramsey's 1928 theory of lawhood, glossed as above, because of its success in explaining some facts about laws of nature. (1) It explains why lawhood is not just a matter of the generality, syntactically or semantically defined, of a single sentence. It may happen that two true sentences are alike general, but one is a law of nature and the other is not. That can happen because the first does, and the second does not, fit together with other truths to make a best system. (2) It explains why lawhood is a contingent property. A generalization may be true as a law at one world, and true but not as a law at another, because the first world but not the second provides other truths with which it makes a best system. (3) It therefore explains how we can know by exhausting the instances that a generalization—say, Bode's 'Law'—is true, but not yet know if it is a law. (4) It explains why *being* a law is not the same as being regarded as a law—being projected, and so forth—and not the same as being regarded as a law and also being true. It allows there to be laws of which we have no inkling. (5) It explains why we have reason to take the theorems of well-established scientific theories provisionally as laws. Our scientific theorizing is an attempt to approximate, as best we can, the true deductive systems with the best combination of simplicity and strength. (6) It explains why lawhood has seemed a rather vague and difficult concept: our standards of simplicity and strength, and of the proper balance between them, are only roughly fixed. That may or may not matter. We may hope, or take as an item of faith, that our world is one where certain true deductive systems come out as best, and certain generalizations come out as laws, by *any* remotely reasonable standards—but we might be unlucky.

On the working hypothesis that the laws of a world are the generalizations that fit into the best deductive systems true there, we can also say that the laws are generalizations which (given suitable companions) are highly informative about that world in a simple way. Such generalizations are important to us. It makes a big difference to the character of a world which generalizations enjoy the status of lawhood there. Therefore similarity and difference of worlds in respect of their laws is an important respect of similarity and difference, contributing weightily to overall similarity and difference. Since a difference in laws would be a big difference between worlds, we can expect that worlds with the same laws as a world *i* will tend to be closer to *i* than worlds at which the laws of *i* hold only as accidental generalizations, or are violated, or—worse still—are replaced by contrary laws. In other words, the laws of *i* will hold throughout many of the spheres around *i*, and thus will tend to be cotenable with counterfactual suppositions. That is so simply because laws are especially important to us, compared with particular facts or true generalizations that are not laws.

Though similarities or differences in laws have some tendency to outweigh differences or similarities in particular facts, I do not think they invariably do so. Suppose that the laws prevailing at a world *i* are deterministic, as we used to think the laws of our own world were. Suppose a certain roulette wheel in this deterministic world *i* stops on black at a time *t*, and consider the counterfactual antecedent that it stopped on red. What sort of antecedent-worlds are closest to *i*? On the one hand, we have antecedent-worlds where the deterministic laws of *i* hold without exception, but where the wheel is determined to stop on red by particular facts different from those of *i*. Since the laws are deterministic, the particular facts must be different at all times before *t*, no matter how far back. (Nor can we assume that the differences of particular fact diminish as we go back in time. Assume for the sake of argument that *i* and its laws are such that any antecedent-world where the laws hold without exception differs more and more from *i* as we go back.) On the other hand, we have antecedent-worlds that are exactly like *i* until *t* or shortly before; where the laws of *i* hold *almost* without exception; but where a small, localized, inconspicuous miracle at *t* or just before permits the wheel to stop on red in violation of the laws. Laws are very important, but great masses of particular fact count for something too; and a localized violation is not the most serious sort of difference of law. The violated deterministic law has presumably not been replaced by a contrary law. Indeed, a version of the violated law, complicated and weakened by a clause to permit the one exception, may still be simple and strong enough to survive as a law. Therefore some of the antecedent-worlds where the law is violated may be closer to *i* than any of the ones where the particular facts are different at all times before *t*. At least, this seems plausible enough to deter me from decreeing the opposite. I therefore proceed on the assumption that the preeminence of laws of nature among cotenable factual premises is a matter only of degree.

My example of the deterministic roulette wheel raises a problem for me: what about differences of particular fact at times *after t*? Among the antecedent worlds I prefer—those where the wheel stops on red by a minor miracle and the particular facts are just as they are at *i* until *t* or shortly before—there are two sorts. There are some where the deterministic laws of *i* are unviolated after *t* and the particular facts after *t* diverge more and more from those of *i*. (I now assume that the deterministic laws are deterministic both forward and backward, so that they do not permit a reconvergence.) There are others where a second minor miracle occurs just after *t*, erasing all traces of the first miracle, so that we have two violations of law instead of one but the particular facts from that time on are just as they are at *i*. If I have decided that a small miracle *before t* makes less of a difference from *i* than a big difference of particular fact at all times *before t*, then why do I not also think that a small miracle *after t* makes less of a difference from *i* than a big difference of particular fact at all times *after t*? That is not what I do think: the worlds with no second miracle and divergence must be regarded as closer, since I certainly think it true (at *i*) that if the wheel had stopped on red at *t*, all sorts of particular facts afterward would have been otherwise than they are at *i*. The stopping on red would have plenty of traces and consequences from that time on.

Perhaps it is just brute fact that we put more weight on earlier similarities of particular fact than on later ones. Divergence of particular fact throughout the past might make more of a difference than a small violation of law, but a small violation of law might make more of a difference than divergence of particular fact throughout the future. Then the closest antecedent-worlds to *i* would be those with a miracle and with no difference of particular fact before *t*, but with no miracle and with divergence of particular fact after *t*. Such discrimination between the two directions of time seems anthropocentric; but we are understandably given to just such anthropocentric discrimination, and it would be no surprise if it turns out to infect our standards of comparative similarity and our truth conditions for counterfactuals.

But perhaps my standards are less discriminatory than they seem. For some reason—something to do with the *de facto* or nomological asymmetries of time that prevail at *i* if *i* is a world something like ours—it seems to take less of a miracle to give us an antecedent-world exactly like *i* in the past than it does to give us one exactly like *i* in the future. For the first, all we need is one little miraculous shove, applied to the wheel at the right moment. For the second, we need much more. All kinds of traces of the wheel's having stopped on red must be falsified. The rest position of the wheel; the distribution of light, heat, and sound in the vicinity; the memories of the spectators—all must be changed to bring about a reconvergence of particular fact between the antecedent-world and *i*. One shove will not do it; many of the laws of *i* must be violated in many ways at many places. Small wonder if the closest antecedent-worlds to *i* are worlds where the particular facts before *t* are preserved at the cost of a small miracle, but the particular facts after *t* are not preserved at the cost of a bigger, more complicated miracle.

3.4 Stalnaker's Theory

The previous theory closest to mine is not any sort of metalinguistic theory, but rather the theory of conditionals put forth by Robert Stalnaker and developed formally by Stalnaker and Richmond Thomason.* According to Stalnaker's theory, a counterfactual $\phi \mathrel{\Box\!\!\rightarrow} \psi$ (he writes it as $\phi > \psi$) is true at a world *i* if and only if either (1) ϕ is true at no world accessible from *i* (the vacuous case), or (2) ψ is true at the ϕ-world closest to *i*.

Stalnaker's theory depends for its success not only on the Limit Assumption that there never are closer and closer ϕ-worlds to *i* without end, but also on a stronger assumption: that there never are two equally close closest ϕ-worlds to *i*, but rather (if ϕ is true at any world accessible from *i*) there is exactly *one* closest ϕ-world. Otherwise there would be no such thing as *the* closest ϕ-world to *i*, and counterfactuals that certainly ought to be true—say, $\phi \mathrel{\Box\!\!\rightarrow} \phi$—would turn out false.

Stalnaker's formal apparatus consists of three things. The first is an *accessibility relation.* Only those worlds that are accessible from *i* need be considered in determining the truth value at *i* of a counterfactual, given the truth values at all worlds of the antecedent and consequent. The second (and principal) one is a *selection function f.* Given any antecedent ϕ and world *i* such that some ϕ-world is accessible from *i*, *f* picks out a single world $f(\phi, i)$: one of the ϕ-worlds accessible from *i*, regarded as the one closest to *i*. The third item is the *absurd world* where everything whatever is true; it is the value of $f(\phi, i)$ if, but only if, there is no ϕ-world accessible from *i*.

Two further formal constraints are imposed, without which we could not regard *f* as making a selection based on comparative similarity. (1) Whenever *i* itself is a ϕ-world, $f(\phi, i)$ is *i*. (2) Whenever ψ holds at $f(\phi, i)$ and ϕ holds at $f(\psi, i)$, $f(\phi, i)$ and $f(\psi, i)$ are the same world.

* Stalnaker, 'A Theory of Conditionals'; Stalnaker and Thomason, 'A Semantic Analysis of Conditional Logic', *Theoria* **36** (1970): 23–42; and Thomason, 'A Fitch-Style Formulation of Conditional Logic', *Logique et Analyse* **52** (1970): 397–412.