The Logical Basis of Phylogenetic Analysis

James S. Farris

Table of Contents

Introduction

Phylogeneticists hold that the study of phylogeny ought to be an empirical science, that putative synapomorphies provide evidence on genealogical relationship, and that (aside possibly from direct observation of descent) those synapomorphies constitute the only available evidence on genealogy. Opponents of phylogenetic systematics maintain variously that genealogies cannot (aside from direct observation) be studied empirically, that synapomorphies are not evidence of kinship because of the possibility of homoplasy, or that raw similarities also provide evidence on genealogy. Most phylogeneticists recognize that inferring genealogy rests on the principle of parsimony, that is, choosing genealogical hypotheses so as to minimize requirements for ad hoc hypotheses of homoplasy. But other criteria as well have been proposed for phylogenetic analysis, and some workers believe that parsimony is unnecessary for that purpose. Others contend* that that principle is not truly "parsimonious," or that its application depends crucially on the false supposition that homoplasy is rare in evolution. Authors of all these criticisms have in common the view that phylogenetic systematics as it is now practiced may be dismissed as futile or at best defective. Phylogeneticists must refute that view,

but accomplishing that goal seems complicated both by the apparent multiplicity of phylogenetic methods and by the diversity of the objections. The methods collectively leave uncertain just what should be defended. Some of the criticisms seem to rule out the possibility of empirically studying phylogeny; some seem to question only parsimony, and some would discard synapomorphy in favor of raw similarity as a guide to kinship. I shall show here that the complexity of this problem is superficial. An analysis of parsimony will not only provide a resolution of the objections to that criterion, but will supply as well an understanding of the relationship of genealogical hypotheses to evidence, and with it a means of deciding among methods of phylogenetic inference.

Ad **Hoc Hypotheses**

I share Popper's disdain for arguing definitions as such, but it is important to make intended meanings clear, and so I shall first dismiss terminological objections to the parsimony criterion. These all come to the idea that parsimonious phylogenetic reconstructions are so primarily by misnomer: the word might equally well refer to any of several other qualities. The meanings of "parsimony" would surely take volumes to discuss, but doing so would be quite pointless. Whether the word is used in the same way by all has no bearing on whether the phylogenetic usage names a desirable quality. I shall use the term in the sense I have already mentioned: most parsimonious genealogical hypotheses are those that minimize requirements for ad hoc hypotheses of homoplasy. If minimizing ad hoc hypotheses is not the only connotation of "parsimony" in general usage, at least it is scarcely novel. Both Hennig (1966) and Wiley (1975) have advanced ideas closely related to my usage. Hennig defends phylogenetic analysis on the grounds of his auxiliary principle, which states that homology should be presumed in the absence of evidence to the contrary. This amounts to the precept that homoplasy ought not be postulated beyond necessity, that is to say parsimony. Wiley discusses parsimony in a Popperian context, characterizing most parsimonious genealogies as those that are least falsified on available evidence. In his treatment, contradictory character distributions provide putative falsifiers of genealogies. As I shall discuss below, any such falsifier engenders a requirement for an ad hoc hypothesis of homoplasy to defend the genealogy. Wiley's concept is then equivalent to mine.

Cartmill (1981) has effectively objected to that last equivalence, claiming that neither phylogenetic analysis nor parsimony can be scientific in Popper's sense. His argument is superficially technical, but his principal conclusion is in fact based on a terminological confusion; and so I shall discuss his ideas here.

Cartmill offers a wide variety of criticisms of parsimony, but most of them have little to do with whether its use is Popperian, and even as objections to parsimony as such, his conclusions are reached primarily by ignoring relevant considerations. He maintains that parsimony cannot be feasibly used to reconstruct genealogy, on the grounds that the attempt would require evaluating a prohibitively great number of alternative trees. He does mention that there are algorithms for circumventing this problem, but claims that all such methods utilize restrictive assumptions (such as the Camin-Sokal method, discussed later). He thus conveniently ignores unrestricted parsimony procedures, such as the Wagner method of Farris (1970). But he glosses over as well a point of more general significance. Scientific inquiry, in Popper's view, as well as those of others, requires only that it be possible to judge competing hypotheses on the basis of observation. There is no demand that the best possible hypothesis be immediately identified. There is, moreover, a strong element of hypocrisy in this part of Cartmill's argument, for this objection amounts just to the claim that it is difficult to be sure that the most parsimonious tree has been found. He does not wish to refrain from drawing evolutionary conclusions himself, however. His own conclusions would, one presumes, be based

on some criteria, although as to what those might be he is rather vague. Unless Cartmill could formulate his own views so as to ensure that finding best conclusions would be a simple task, he would fall prey to his own argument.

Cartmill also argues at length that most parsimonious (Wagner) trees do not determine the features of the most recent common ancestor of their terminal taxa. Why he felt it necessary to do so seems rather a mystery, as the property had long been known (Farris, 1970). He portrays this as a criticism simply by ignoring the way in which such trees do yield inferences on plesiomorphy: through the outgroup criterion. He discusses at similar length a formulation of parsimony by Engelmann and Wiley (1977), which he supposes to differ from parsimony in the present (Wagner) sense. Wiley (pers. comm.; see also Wiley, 1981) denies this difference; it appears that Cartmill simply misunderstood their procedure. To the extent that his discussion of this method is accurate, it seems to lead again just to the same conclusion: that parsimony does not determine ancestral features ("morphotypes," as he quaintly puts it) without outside evidence. His own favorite source of such evidence is paleontological information. This high esteem for fossils, however, has no force as a criticism of parsimony methods. A fossil, after all, might perfectly well be used as an outgroup. Why Cartmill supposed any of this is either a new idea or a criticism of phylogenetic analysis is, again, a mystery. As much can be found in Hennig (1966). None of this does anything to support the accusation that phylogenetic methods are non-Popperian, for Cartmill does not claim that hypotheses of plesiomorphy are untestable by paleontological means.

Cartmill does offer two arguments that at least seem to address directly the Popperian falsifiability of genealogical hypotheses. In one, he considers the premise that unparsimonious arrangements are necessarily false, rejecting the use of that idea in a Popperian falsification argument as both unrealistic and infeasible. Inasmuch as that premise is entirely Cartmill's own creation, resembling nothing that any phylogeneticist has ever suggested, this part of his reasoning is quite irrelevant to phylogenetic practice. One can only suppose that his intention here was to "criticize" phylogenetic systematics by attacking a straw man.

In the other argument, Cartmill cites Gaffney (1979) to the effect that character distributions are falsifiers of genealogical hypotheses, and that it is possible that every conceivable genealogy will be falsified at least once. From the first of these admissions he "deduces" that Gaffney must have relied on the "theorem" that any genealogy contradicted by a character distribution is false. Cartmill then reasons: some genealogy must be true. "Gaffney's" theorem, together with a falsifier for every genealogy, implies that every genealogy is false. Therefore, Gaffney's claim, that character distributions are falsifiers, is false.

Cartmill's argument rests directly and entirely on a misrepresentation of the Popperian meaning of "falsifier:" a test statement that, if true, allows a hypothesis to be rejected. There is a great difference between "falsify" in Popper's sense and "prove false." The relationship between a theory and its falsifiers is purely logical; Popper never claimed that proof of falsity could literally be achieved empirically. "Observing" a falsifier of a theory does not prove that the theory is false; it simply implies that either the theory or the observation is erroneous. It is then seen that the only implication that can be derived from falsification of every genealogy is that some of the falsifiers are errors-homoplasies. It is thus seen as well that Cartmill's "syllogism" is nothing other than an equivocation.

So much for the claim that characters cannot be Popperian falsifiers, but is phylogenetic parsimony Popperian? Cartmill admits that Phylogeneticists hold that the least falsified genealogy is to be preferred. The reason for this preference is that each falsifier of any accepted genealogy imposes a requirement for an ad hoc hypothesis to dispose of the falsifier. According to Popper—as Cartmill also cites—ad hoc hypotheses must

be minimized in scientific investigation. Cartmill never attempts to argue that conflicts between characters and genealogies do not require hypotheses of homoplasy, and so none of his claims can serve to question the connection between parsimony and Popper's ideas.

Parsimony and Synapomorphy

Matters of usage having been disposed of, we may turn to a more substantial question: whether parsimony in phylogenetic analysis is desirable. Part of the reason for thinking that it might not be is the suspicion—Hennig and Wiley notwithstanding—that it may be possible to dispense with parsimony in Hennigian methods. The objection that parsimony requires rarity of homoplasy in evolution is usually taken to be just that: a criticism of parsimony. It might seem that the problem posed by that objection could be avoided simply by using some other criterion for phylogenetic analysis. Some quite non-phylogenetic proposals, such as grouping according to raw similarity, have been made along those lines, and I shall discuss those eventually. Of more immediate interest is the question of whether grouping by putative synapomorphy can do without the parsimony criterion.

The view that parsimony is unnecessary to phylogenetic methods is well exemplified by Watrous and Wheeler (1981:9), according to whom, "If the character is incongruent with other characters, we have three alternatives. We can select the most parsimonious arrangement. . . . We can reinterpret the homologies of the characters. . . . Or, we can weight characters to resolve conflicts." If parsimony is an alternative to weighting or to removing conflicts, it would seem, then parsimony would not be needed if either course were taken. In fact, neither process can obviate minimizing requirements for ad hoc hypotheses of homoplasy.

In any case in which characters seem to offer conflicting evidence on the placement of a taxon, selecting one placement requires that the characters disputing that choice be dismissed as evidence. That dismissal is defended by concluding that those characters are homoplasious. Sometimes extrinsic reasons can be found to justify the conclusion of homoplasy. Scrutiny of a structure may indicate that what had been regarded as a single feature is instead the union of distinct qualities. Sometimes conflicts can be removed by changing hypotheses of plesiomorphy, and surveying the distribution of a feature among taxa may provide support for such a reinterpretation. If a conflicting character survives all attempts to remove it by searching for such evidence, then the conclusion of homoplasy in that character, required by selecting a placement, satisfies the usual definition of an ad hoc hypothesis. It is required to defend the genealogical hypothesis chosen, but it is not supported by any evidence separate from that for the genealogy itself. If external evidence favors the interpretation of homoplasy, however, that hypothesis is not ad hoc. Usually a careful worker will have studied his characters closely before attempting to fit them into a synapomorphy scheme, with the consequence that most cases of homoplasy concluded from a phylogenetic analysis will comprise ad hoc hypotheses. In any case, the reinterpretation suggested by Watrous and Wheeler cannot be separate from the aim of minimizing ad hoc hypotheses of homoplasy. If the reinterpretation is based on evidence, its effect is precisely to remove such ad hoc hypotheses whenever possible.

In portraying weighting as an alternative to parsimony, Watrous and Wheeler apparently intended to equate the parsimony criterion with simple counting of equally weighted homoplasies. That usage reflects both a lack of familiarity with the way in which parsimony has long been used by other Phylogeneticists and a misunderstanding of the nature of character weighting. Kluge and Farris (1969), whom, incidentally, Watrous and Wheeler cite, for example, use character weighting in conjunction with parsimony,

evaluating the latter by finding a weighted total of homoplasies. The reason for doing so is plain enough. Genealogies are selected to avoid requirements for ad hoc hypotheses of homoplasy because characters imposing such requirements are regarded as evidence favoring alternative genealogies over the one selected. In the absence of any convincing reason for doing otherwise, the characters of a study are often treated in practice as if they all provided equally cogent evidence on phylogenetic relationship. No one supposes, however, that characters in general all deserve the same weight—that they all yield equally strong evidence. Drawing conclusions despite conflicting evidence requires that some evidence be dismissed as homoplasy. It is surely preferable to dismiss weaker evidence in deference to stronger. A decision reached by weighting characters, at any rate, can hardly rest on a basis different from parsimony. The effect of giving a character a weight X is just to proceed as if the data included X independent characters all showing the same distribution of states. Now suppose that many independent characters support one placement of a taxon, while just one supports an alternative placement. Possible reinterpretations aside, if the characters are weighted equally, weight of evidence favors the first placement. If the data were different and the counts reversed, the second placement would then be favored. If character weights were not all equal, either placement might be supported by the greater weight of evidence, depending on the character weights. The process of selecting a placement would be the same whether weight of evidence were reflected by counts of equal weights or by sums of differing ones. In either case the decision is made by accepting the stronger body of evidence over the weaker, and ad hoc hypotheses of homoplasy are required to the extent that evidence must be dismissed in order to defend the conclusion.

Watrous and Wheeler's suggestion that parsimony is needed only when characters conflict has the implication that a set of congruent characters can be analyzed without avoiding ad hoc hypotheses of homoplasy. A similar idea would appear to underlie advocacy by Estabrook and others (reviewed by Farris and Kluge, 1979) of techniques ("clique" methods) that "resolve" character conflicts by discarding as many characters as necessary so that those surviving (the clique) are mutually congruent. The surviving characters are then used to construct a tree. Proponents of such methods maintain that the tree so arrived at rests on a basis different from parsimony.

The character selection process itself may well have a distinctive premise, a possibility that I shall discuss below. To claim that the interpretation of the characters selected rests on a basis other than parsimony, however, seems not to be defensible. The tree constructed from a suite of congruent characters by a clique method is chosen to avoid homoplasy in any of the characters, the possibility of doing so being ensured by the selection (selection aside, Watrous and Wheeler proceed likewise). It seems accurate, then, to describe that construction as minimizing requirements for ad hoc hypotheses of homoplasy for the characters within the congruent suite, but, more particularly, there seems to be no other sensible rationale for the construction. No one seems to have suggested any such principle, aside from the obvious: that if the characters were free of homoplasy (were "true," as it is often put), then the tree would follow. But the characters comprising a congruent suite are hardly observed to be free of homoplasy. At the most it might be said that the selected characters seem to suggest no genealogy other than the obvious one.

Of course it is what data suggest, or how they do it, that is at issue. If a suite of congruent characters is interpreted by avoiding unnecessary postulates of homoplasy, then the interpretation embodies parsimony. But the only apparent motivation for concentrating just on congruent characters is to avoid reliance on parsimony. That avoidance would seem sensible only on the supposition that parsimony is ill founded, and the only apparent reason for that supposition is the charge that parsimony depends crucially on unrealistic assumptions about nature. If that charge means anything at all, it must

mean that taking conditions of nature realistically into account would lead to preference for **a** less parsimonious arrangement over **a** more parsimonious one. But if that charge were correct, then it would be—to say the least—less than obvious why the implications of those natural conditions would be expected to change simply because any characters incongruent with those chosen had been ignored.

If avoiding ad hoc hypotheses of homoplasy is unjustified, then neither Watrous and Wheeler nor clique advocates are entitled to the inferences on phylogeny that they draw; but the significance of parsimony for Hennigian methods is much more general than that. Watrous and Wheeler probably thought that they had no need of anything so questionable as parsimony, because they were simply applying Hennig's well-established principle of grouping according to synapomorphy. Just how did that principle come to be well established? It is usually explained by taking note of the logical relationship between monophyletic groups and true synapomorphies, but that leaves open the question of how genealogies are related to observed features. It might well be questioned whether the logical construct can legitimately be extended into a principle to guide interpretation of available characters. That question has in fact often been raised, **and** almost always in the form of the suggestion that putative synapomorphies are not evidence of kinship because they might well be homoplasies. Hennig's (1966) own reply to that objection was his auxiliary principle, which, as I have already observed, is **a** formulation of the parsimony criterion.

Hennig's defense of the synapomorphy principle by recourse to parsimony is not accidental, but necessary. The analytic relationship of correct synapomorphies to phylogeny is just that a property that evolved once and is never lost must characterize a monophyletic group. Synapomorphies are converted into a genealogy, that is, by identifying the tree that allows a unique origin for each derived condition. A phylogeny based on observed features is parsimonious to the degree that it avoids requirements for homoplasies—multiple origins of like features. Secondary plesiomorphies aside, a plesiomorphic trait will already have a single origin at the root of the putative tree, so that the effect of parsimony is precisely to provide unique derivations wherever possible. (Secondary plesiomorphies, being a kind of apomorphy, are treated likewise.) Grouping by synapomorphy would thus have to behave like parsimony, but further, the latter applies to actual traits, whereas the logic of true synapomorphies does not. Superficially, the use of the synapomorphy principle in phylogenetic inference seems to be just a consequence of the logical connection between true synapomorphies and genealogies, but it cannot be just that, as the condition of that logic—that the traits are indeed synapomorphies—need not be met. Grouping by putative synapomorphy is instead **a** consequence of the parsimony criterion.

Abundance of Homoplasy

That the synapomorphy principle is based on parsimony naturally brings to attention the role that homoplasy has played in objections to both. There are two main varieties of the position that use of the parsimony criterion depends crucially on the supposition that homoplasy is rare in evolution. In the first, the observation that requirements of homoplasy are minimized is taken as prima fecie evidence that the supposition is needed. In the second, the claim is advanced in conjunction with some more elaborate, often statistical, argument. The conclusion from the first kind of reasoning is quite general, while that from the second is necessarily limited by the premises of the argument employed. If the first kind of criticism were correct, there would be little point to considering arguments of the second sort. I shall thus first point out why the first type of objection rests on **a** fallacy.

To evaluate the claim that an inference procedure that minimizes something must ipso facto presuppose that the quality minimized is rare, it is useful to consider a common application of statistics. In normal regression analysis, a regression line is calculated from a sample of points so as to minimize residual variation around the line, and the residual variation is then used to estimate the parametric residual variance. Plainly the choice of line has the effect of minimizing the estimate of the residual variance, but one rarely hears this procedure criticized as presupposing that the parametric residual variance is small. Indeed, it is known from normal statistical theory that the least squares line is the best point estimate of the parametric regression line, whether the residual variance is small or not. The argument that the parsimony criterion must presume rarity of homoplasy just because it minimizes required homoplasy is thus at best incomplete. That reasoning presumes a general connection between minimization and supposition of minimality, but it is now plain that no such general connection exists. Any successful criticism of phylogenetic parsimony would have to include more specific premises.

The same conclusions can readily be reached in a specifically phylogenetic context. Suppose that for three terminal taxa A, B, C there are 10 putative synapomorphies of A + B and 1 putative apomorphy shared by B and C. We assume for simplicity of discussion that the characters are independent and all of equal weight, and that attempts to find evidence to support changes in the data have already failed. Parsimony then leads to preference for ((A, B), C) over alternative groupings. We will be interested in whether abundance of homoplasy leads to preference for some other grouping. If it does not, then the claim that parsimony presupposes rarity of homoplasy is at best not generally true. It will be useful to examine first the relationship between the genealogical truth of the grouping and homoplasy in particular characters, then to consider the effect of general abundance of homoplasy.

It is plain that the grouping ((B, C), A) is genealogically correct if the one B + C character is in fact a synapomorphy, and that ((A, B), C) is instead correct if the A + B characters are synapomorphies. Truth of the latter grouping does not require, however, that all 10 of the putative synapomorphies of A + B be accurate homologies. If just one of those characters were truly a synapomorphy, while all the other characters in the data were in fact parallelisms, the genealogy would necessarily be ((A, B), C). That A and B share a common ancestor unique to them, in other words, does not logically require that every feature shared by A and B was inherited from that ancestor. In the extreme, if all the characters were parallelisms, this would not imply that ((A, B), C) is genealogically false. Under those circumstances the data would simply leave the question of the truth of that (or any other) grouping entirely open.

The relationship between characters and genealogies thus shows a kind of asymmetry. Genealogy ((A, B), C) requires that the B +C character be homoplasious, but requires nothing at all concerning the A + B characters. The genealogy can be true whether the conforming characters are homoplasious or not. One kind of objection to phylogenetic parsimony runs that ad hoc hypotheses are indeed to be minimized, but this does not mean minimizing homoplasies, because a genealogy also requires ad hoc hypotheses of homology concerning the characters that conform to it. It is seen that such is not the case. Only characters conflicting with a genealogy lead to requirements for ad hoc hypotheses, and so the only ad hoc hypotheses needed to defend a genealogy are hypotheses of homoplasy.

The sensitivity of inference by parsimony to rarity of homoplasy is readily deduced from these observations. If homoplasy is indeed rare, it is quite likely with these characters that ((A, B), C) is the correct genealogy. In order for that grouping to be false, it would be required at least that all 10 of the A + B characters be homoplasious. As these characters are supposed to be independent, the coincidental occurrence of homo-

plasy in all 10 should be quite unlikely. Suppose then that homoplasy is so abundant that only one of the characters escapes its effects. That one character might equally well be any of the 11 in the data, and if it is any one of the 10 A+ B characters the parsimonious grouping is correct. That grouping is thus a much better bet than is ((B, C), A). At the extreme, as has already been seen, if homoplasy is universal, the characters imply nothing about the genealogy. In that case the parsimonious grouping is no better founded than is any other, but then neither is it any worse founded.

It seems that no degree of abundance of homoplasy is by itself sufficient to defend choice of a less parsimonious genealogy over a more parsimonious one. That abundance can diminish only the strength of preference for the parsimonious arrangement; it can never shift the preference to a different scheme. In this the relationship of abundance of homoplasy to choice of genealogical hypothesis is quite like that between residual variance and choice of regression line. Large residual variance expands the confidence interval about the line, or weakens the degree to which the least squares line is to be preferred over nearby lines, but it cannot by itself lead to selection of some other line that fits the data even worse.

Stochastic Models

The supposition of abundance of homoplasy by itself offers no grounds for preferring unparsimonious arrangements, but it is easy enough to arrive at that preference by resorting to other premises. Felsenstein (1973, 1978, 1979) has objected to parsimony on statistical grounds. He suggests (as others have) that genealogies ought to be inferred by statistical estimation procedures. In his approach he devises stochastic models of evolution, then applies the principle of maximum likelihood, choosing the genealogy that would assign highest probability to the observed data if the model were true. With the models that he has investigated, it develops that the maximum likelihood tree is most parsimonious when rates of change of characters are very small, under which circumstances the models would also predict very little homoplasy. He concludes from this that parsimony requires rarity of homoplasy. In his 1978 paper he discusses a model according to which both parsimony and clique methods would be certain to yield an incorrect genealogy if a large enough random sample of independent characters were obtained. He contends that maximum likelihood estimation under the same conditions would yield the correct tree, as that estimate possesses the statistical property termed consistency. That last is the logical property that if an indefinitely great number of independent characters were sampled at random from the distribution specified by the model, then the estimate would converge to the parameter of the model, the hypothetical true tree.

Whatever else may be true of parsimony, it is certainly intended as a criterion to be employed in systematic practice. One is then immediately suspicious of a criticism that involves statistically random sampling of indefinitely many characters from a single statistical population. Organisms can no more have an indefinitely great number of characters than they can have an indefinitely great quantity of DNA. Even granting that the number of potentially usable characters is large, moreover, it hardly seems accurate to describe systematic research as random sampling of unitary characters that lie waiting to be selected. For morphological traits, at least, characters might better be considered to be synthesized by the systematist as summaries of lower level observations on the differences among organisms—and even those observations are unlikely to be made at random. Felsenstein's argument from consistency, furthermore, depends on still further suppositions. In the two more recent papers, characters are abstracted as binary qualities whose states may arise repeatedly and indistinguishably. A morphologist might well be willing to concede that traits in some sense the same can arise repeatedly, **but** he **might**

well also expect that independent derivations would often be discovered to be such by closer inspection, in which case the character would be coded as having more than two states. This is more than just a question of coding. In morphological studies, sameness (homology) of traits of different organisms is a hypothesis tentatively arrived at by detailed inspection, and so traits showing signs of difference are unlikely to-be regarded as the same from the start. In Felsenstein's abstraction, homology of traits is simply stipulated as part of the model, and there is no way for evidence to be brought to bear on it. In all the models it is supposed that all characters are drawn from the same statistical population, a condition ensured by supposing in turn that all characters have the same chance of changing in any given time interval. In his first paper rates of change are also presumed to be homogeneous over time and among phyletic lines as well. A biologist might consider it more likely that (for example) horseshoe crabs have evolved more slowly than insects, that tooth counts change more frequently in fishes than in mammals, and that paired appendages generally do the opposite. Indeed, the general success of hierarchic summaries of characters of organisms seems to suggest that each feature undergoes evolution in only a restricted part of the evolutionary tree, and that different portions of the tree have different suites of varying features. That is, after all, what is implied by the observation that the synapomorphies of tetrapods involve different traits than do those of hymenopterans.

There is more to these observations than just lack of realism in the models as such. Those peculiar suppositions are in fact crucial to Felsenstein's argument. The quantity consistently estimated in his treatment exists formally just because of those premises. His methods estimate the average probability of change in each branch of the evolutionary tree. As all characters are presumed to bear on that average in just the same way, the estimated averages should converge to the true averages (if those existed) as the number of randomly sampled characters becomes large. But without homogeneity of characters or random sampling, the conclusion does not follow. But there is still more. Consistency is grounded in a reasonable idea, in its way. Other things being equal, it would be desirable if an estimate would converge to the correct value as the amount of evidence increased. Even here, however, Felsenstein has used an assumption, for the only source of additional evidence in his scheme is further random sampling of characters. A biologist might consider that a tree would become well founded if a large number of fossils were discovered, replacing the separations between the original terminals with a near-continuum. But if new evidence were introduced in this way, there would be no guarantee that the branch averages—which now themselves would become numerous—would be consistently estimated. Felsenstein has arrived at his conclusions on consistency simply by ignoring considerations that would be relevant in practice. It would appear, in fact, that he chose his models just to facilitate a formal consistency argument, any considerations of realism being for this purpose deliberately ignored. His models are evolutionary only in a purely formal sense, and further, insofar as they also in effect model systematic research, they do not seem to model even the right sort of process.

Felsenstein does not try to defend his models as realistic. His attitude on their purpose seems to be instead that "if a method behaves poorly in this simple model framework, this calls into question its use on real characters" (1981:184), or perhaps that "if phylogenetic inference is to be a science, we must consider its methods guilty until proven innocent" (1978:409). The first is preposterous except on the supposition that reasoning from false premises cannot lead to false conclusions. As for the second: to the extent that these models are intended seriously, they comprise empirical claims on evolution. If science required proof concerning empirical claims in order to draw conclusions, no kind of science would be possible.

Felsenstein nonetheless apparently believed that he had demonstrated that practical

application of parsimony requires rarity of homoplasy, but in fact such is hardly the case. Purely technically, it is less than obvious that he has identified the most general conditions under which maximum likelihood and parsimonious estimates coincide; indeed, he seems to have concentrated on finding cases in which they do not. But other aspects of this issue are far more important. The dependence of parsimony on rarity of homoplasy is in Felsenstein's analysis a consequence of his models. These models, as he is well aware, comprise "strong assumptions about the biological situation" (1978:403). If those assumptions do not apply to real cases, then, so far as Felsenstein can show, the criticism of parsimony need not apply to real cases either. But, again, Felsenstein does not maintain that the assumptions of his models are realistic. He has not shown that abundance of homoplasy implies preference for unparsimonious genealogies. Instead he has shown at most that if homoplasy were abundant, and if in addition the conditions of his models prevailed in nature, then one should prefer unparsimonious schemes. We have already seen that abundance of homoplasy by itself does not justify departure from parsimony. If Felsenstein's argument offers any reason for that departure, then, that reason would have to rely on the supposition that his models apply to nature. An ironic result indeed. The original criticism of parsimony was that it required an unrealistic assumption about nature. It now seems instead that unparsimonious methods require such assumptions, whereas parsimony does not.

That is not to say that parsimony requires no assumptions at all; it presumes, one might say, that Felsenstein's models are unrealistic. But as that assumption seems generally agreed upon, that is not much of a criticism of parsimony. Felsenstein's arguments from consistency and maximum likelihood have a related drawback. Consistency is a logical relationship between an estimation method and a probability model. In the hypothetical case imagined by Felsenstein, his method would have obtained the right answer, but whether the method would work in practice depends on whether the model is accurate. If it is not, then the consistency of the estimator under the model implies nothing about the accuracy of the inferred tree. The status of a procedure as a maximum likelihood estimator is also bound to the probability model. If the model is false, the ability of a procedure to find the most likely tree under the model implies nothing of how likely the chosen tree might actually be. Likewise, the conclusion that parsimony would arrive at the wrong tree depends on the model, and so the hypothetical analysis implies little about the practical accuracy of parsimony.

One might say, of course, that the model illustrates a potential weakness of parsimony: that criterion will fail if the conditions of the model should happen to be met. And how are we to know that this will not happen? This seems in fact to be the intended substance of Felsenstein's remarks. While admitting that his premises are unrealistic, he rejects that as a criticism of his attack on parsimony, claiming that an objection based on realism "amounts to a confession of ignorance rather than a validation of the inference method in question" (1978:408). A derivation that implies nothing about reality is not much of an improvement on ignorance, of course. To the extent that Felsenstein has a point, then, it seems to be just that parsimony is invalid because we cannot be certain that it will not lead to errors of inference. But there is nothing distinctive about Felsenstein's model in that regard. One may always concoct fantastic circumstances under which scientific conclusions will prove incorrect. It is hardly necessary to resort to mathematical manipulations to produce such fears. One need only imagine that his characters have evolved in just the right way to lead him to a false conclusion. Or, with Descartes, that his perceptions and reason have been systematically and maliciously distorted by a demon. None of these possibilities can be disproved, but it hardly matters. There is likewise no reason for accepting any of them, and so collectively they amount to no more than the abstract possibility that a conclusion might be wrong. No phylogeneticist—or any scientist—would dispute that anyway, and so such "objections" are

entirely empty. That thinking provides no means of improving either conclusions or methods, but instead offers, if anything, a rejection of all conclusions that cannot be established with certainty. If that attitude were taken seriously, no scientific conclusions whatever could be drawn.

Explanatory Power

A number of authors, myself among them (Farris, 1973, 1977, 1978), have used statistical arguments to defend parsimony, using, of course, different models from Felsenstein's. Felsenstein has objected to such derivations on grounds of statistical consistency: as before, parsimonious reconstructions are not consistent under his models. That is no more than an equivocation, as the models differ, and consistency is a relationship between method and model. But I do not mean by this to defend those favorable derivations, for my own models, if perhaps not quite so fantastic as Felsenstein's, are nonetheless like the latter in comprising uncorroborated (and no doubt false) claims on evolution. If reasoning from unsubstantiated suppositions cannot legitimately question parsimony, then neither can it properly bolster that criterion. The statistical approach to phylogenetic inference was wrong from the start, for it rests on the idea that to study phylogeny at all, one must first know in great detail how evolution has proceeded. That cannot very well be the way in which scientific knowledge is obtained. What we know of evolution must have been learned by other means. Those means, I suggest, can be no other than that phylogenetic theories are chosen, just as any scientific theory is, for their ability to explain available observations. I shall thus concentrate on evaluating proposed methods of phylogenetic analysis on that basis.

As the supposition that homoplasy is abundant is not alone sufficient to justify preference for unparsimonious arrangements, that preference must be based on other premises. Reasons for departing from parsimony are generally put forward as parts of arguments favoring other approaches to phylogenetic inference, and it is productive to consider these ideas in connection with the methods which they are used to rationalize. For this purpose it will be useful to examine just why it is desirable to avoid ad hoc hypotheses of homoplasy in choosing among genealogical hypotheses, and just how genealogies are evaluated through this principle.

That ad hoc hypotheses are to be avoided whenever possible in scientific investigation is, so far as I am aware, not seriously controversial. That course is explicitly recommended by Popper, for example. No one seems inclined to maintain that ad hoc hypotheses are desirable in themselves; at most they are byproducts of conclusions held worthy on other grounds. Nonetheless, I suspect that much of the criticism of the phylogenetic parsimony criterion arises from a failure to appreciate the reasons why ad hoc hypotheses must be avoided. Avoiding them is no less than essential to science itself. Science requires that choice among theories be decided by evidence, and the effect of an ad hoc hypothesis is precisely to dispose of an observation that otherwise would provide evidence against a theory. If such disposals were allowed freely, there could be no effective connection between theory and observation, and the concept of evidence would be meaningless. The requirement that a hypothesis of kinship minimize ad hoc hypotheses of homoplasy is thus no more escapable than the general requirement that any theory should conform to observation; indeed, the one derives from the other.

It will be profitable to analyze conformity of theory to observation in greater detail. There are a number of properties commonly held to characterize a theory that gives a satisfactory account of observations. The theory must first of all provide a description of what is known, else it would serve little purpose. As Sober (1975) puts it, theories serve to make experience redundant. But not all descriptions are equally useful. Good theories describe in terms of a coherent framework, so that experience becomes com-

prehensible; in short they are explanatory. Explanations in turn may be judged on their ability to cover observations with few boundary conditions, that is, with little extrinsic information. Sober has characterized theories satisfying this goal as most informative, or simplest. All these criteria are interrelated in the case of phylogenetic inference, so that they effectively yield a single criterion of analysis. These connections have already been recognized, and I shall summarize them only briefly. I shall be concerned mostly with analyzing the implications of that one criterion in terms of explanatory power.

I have elsewhere (Farris, 1979, 1980, 1982b) already analyzed the descriptive power of hierarchic schemes. I showed that most parsimonious classifications are descriptively most informative in that they allow character data to be summarized as efficiently as possible. That conclusion has aroused some opposition, as syncretistic taxonomists had been inclined to suppose that grouping according to (possibly weighted) raw similarity gave hierarchies of greatest descriptive power. There seems to be no reason for taking that view seriously, however, as no attempt has been made to derive clustering by raw similiarity from the aim of effective description of character information. A few attempts have been made to justify unparsimonious arrangements on grounds of descriptive efficiency, but (as reviewed by Farris, 1982b) these have all depended on incorrectly calculated examples.

In my treatment I found that a hierarchic classification provides an informative or efficient description of the distribution of a feature to the degree that the feature need occur in the diagnoses of few taxa. The utility of efficient descriptions is precisely that they minimize redundancy. As I have observed before (particularly Farris, 1980), the presence of a feature in the diagnosis of a taxon corresponds to the evolutionary interpretation that the feature arose in the stem species of that taxon. There is thus a direct equivalence between the descriptive utility of a phylogenetic taxon and the genealogical explanation of the common possession of features by members of that group. Sober (1975) has stressed the importance of informativeness of theories, and has developed a characterization of informativeness in terms of simplicity. It is no surprise to find that simplicity is related to parsimony, and Beatty and Fink (1979) have lucidly discussed the connection in terms of Sober's ideas. Sober (1982a) has likewise concluded that phylogenetic parsimony corresponds to simplicity (efficiency, informativeness) of explanation.

In choosing among theories of relationship on the basis of explanatory power, we wish naturally to identify the genealogy that explains as much available observation as possible. In general, deciding the relative explanatory power of competing theories can be a complex task, but it is simplified in the present case by the fact that genealogies provide only a single kind of explanation. A genealogy does not explain by itself why one group acquires a new feature while its sister group retains the ancestral trait, nor does it offer any explanation of why seemingly identical features arise independently in distantly related lineages. (Either sort of phenomenon might of course be explained by a more complex evolutionary theory.) A genealogy is able to explain observed points of similarity among organisms just when it can account for them as identical by virtue of inheritance from a common ancestor. Any feature shared by organisms is so either by reason of common descent or because it is a homoplasy. The explanatory power of a genealogy is consequently measured by the degree to which it can avoid postulating homoplasies.

It is necessary in applying that last observation to distinguish between homoplasies postulated by the genealogy and those that are concluded for other reasons. A structure common to two organisms might be thought to be a homoplasy on grounds extrinsic to the genealogy, as in the reinterpretations discussed earlier. Such a conclusion would amount to specifying that the structure is not a point of heritable similarity. A genealogy would not explain such a similarity, but that would be no grounds for criticizing the

genealogy. Rather, the extrinsic conclusion would make the feature irrelevant to evaluating genealogies by effectively stipulating that there is nothing to be explained. The same would hold true for any trait that is known not to be heritable, such as purely phenotypic variations. The explanatory power of a genealogy is consequently diminished only when the hypothesis of kinship requires ad hoc hypotheses of homoplasy.

By analogy with the abundance of homoplasy argument, it might be objected that seeking a genealogical explanation of similarities is pointless, inasmuch as most similarities are likely to be homoplasies anyway. If homoplasy were universal, that point might well hold. It seems unlikely, however, that homoplasy is universal. It is seldom maintained that segmented appendages have arisen independently in each species of insect. Universality of homoplasy would imply in the extreme that organisms do not generally resemble their parents, a proposition that seems at best contrary to experience. That the character distributions of organisms generally correspond to a hierarchic pattern, furthermore, seems comprehensible only on the view that the character patterns reflect a hierarchy of inheritance. Indeed, the recognized organic hierarchy was one of the chief lines of evidence for Darwin's theory of descent with modification. The idea that homoplasy is abundant is not usually intended in such extreme form, of course, Usually it is meant just to suggest that there is room for doubt concerning whether a shared feature is a homology or a homoplasy. Under those circumstances, however, genealogies retain explanatory power. More to the point, the explanatory power of alternative genealogies is still related to their requirements for homoplasies. Suppose that one genealogy can explain a particular point of similarity in terms of inheritance, while a second hypothesis of kinship cannot do so. If that point of similarity is in fact a homoplasy, the similarity is irrelevant to evaluating genealogical hypotheses, as has already been seen. If the similarity is instead a homology, then only the first genealogy can explain it. If there is any chance that *the* similarity is homologous, the first genealogy is to be preferred.

There is nothing unusual in the relationship of genealogical hypotheses to characters. Scientific theories are generally chosen to conform to data, but it is seldom possible to guarantee that observations are free of errors, and it is no criticism of a theory if it turns out that some of the observations that conform to it are susceptible to error. If a theory does not conform to some observation, however, then the mere suspicion that the observation might be erroneous is not logically adequate to save the theory. Instead the data must be dismissed outright by recourse to an ad hoc hypothesis. Establishing that an observation is erroneous, on the other hand, simply makes it irrelevant to evaluating the theory. The relationship between the explanatory power of genealogies and their requirements for ad hoc hypotheses is likewise characteristic of theories in general. Any observation relevant to evaluating a theory will either conform—and so be explained—or fail to do so, in which case an ad hoc hypothesis is needed to defend the theory. It is generally true that a theory explains relevant observations to the degree that it can be defended against them without recourse to ad hoc hypotheses.

Independence of Hypotheses

Identifying the more explanatory of two alternative hypotheses of kinship is accomplished by finding the total of ad hoc hypotheses of homoplasy required by each. Reckoning those totals will generally involve summing both over separate characters and over observed similarities within characters. Only required ad hoc hypotheses diminish the explanatory power of a putative genealogy. It is thus important to ensure that the homoplasies combined in such totals are logically independent, since otherwise their number need not reflect required ad hoc hypotheses. If two characters were logically or functionally related so that homoplasy in one would imply homoplasy in the other, then

homoplasy in both would be implied by a single ad hoc hypothesis. The "other" homoplasy does not require a further hypothesis, as it is subsumed by the relationship between the characters. This is the principle underlying such common observations as that only independent lines of evidence should be used in evaluating genealogies, and that there is no point to using both number of tarsal segments and twice that number as characters. Phylogeneticists seldom attempt to use logically related characters as separate sources of evidence (although an example of this mistake is discussed by Riggins and Farris, 1982), and so it seems unnecessary to discuss this point further here.

A different sort of interdependence among homoplasies may arise in considering similarities within a single character. Suppose that 20 of the terminal taxa considered show a feature X, and that a putative genealogy distributes these taxa into two distantly related groups A and B of 10 terminals each. There are 100 distinct two-taxon comparisons of members of A with members of B, and each of those similarities in X considered in isolation comprises a homoplasy. Those homoplasies do not constitute independent required hypotheses, however. The genealogy does not require that similarities in X within either group be homoplasies; it is consistent with identity by descent of X within each group. If X is identical by descent in any two members of A, and also in any two members of B, then the A-B similarities are all homoplasies if any one of them is. The genealogy thus requires but a single ad hoc hypothesis of homoplasy. Of course the numbers in the groups do not matter; the same conclusion would follow if they were 15 and 5, or 19 and 1.

Similar reasoning can be extended to more complex examples, but the problem can be analyzed more simply. If a genealogy is consistent with a single origin of a feature, then it can explain all similarities in that feature as identical by descent. A point of similarity in a feature is then required to be a homoplasy only when the feature is required to originate more than once on the genealogy. A hypothesis of homoplasy logically independent of others is thus required precisely when a genealogy requires an additional origin of a feature. The number of logically independent ad hoc hypotheses of homoplasy in a feature required by a genealogy is then just one less than the number of times the feature is required to originate independently. (The lack of a structure might of course be a feature; "origins" should be interpreted broadly, to include losses.)

Lengths

That last observation reduces to the rule that genealogies with greatest explanatory power are just those that minimize the (possibly weighted) total of required independent origins of known features. There is another way of putting that characterization, in terms of length. Each required origin of a feature can be assigned (although not necessarily uniquely) to a particular branch of a putative tree, and the weighted total of the origins in a branch can be regarded as that branch's length. If such lengths are summed over branches of the tree, the result is the total required homoplasy, or the length of the tree. Early work on automatic techniques of parsimony analysis (particularly the Wagner method formulation of Kluge and Farris, 1969) used the length conception of parsimony. That formulation has turned out to be technically very useful and has facilitated considerable progress in methods of analysis. (Basic principles are described by Farris, 1970; for applications of some greatly improved procedures see Mickevich, 1978, 1980; Schuh and Farris, 1981; Mickevich and Farris, 1981). Nonetheless its use was in a way unfortunate, for the length terminology has probably caused more misunderstanding than has any other single aspect of parsimony methods.

The length measure used by Kluge and Farris is coincidentally a familiar mathematical measure of distance in abstract spaces, the Manhattan metric. Once ideas have been reduced to formulae, it is easy to forget where the formulae came from, and to devise

new methods with no logical basis simply by modifying formulae directly. Phylogenetic reconstructions typically infer the features of hypothesized ancestors, so that the length of a branch lying between two nodes of a tree can be regarded as the distance between two points in the space of possible combinations of features. If one notes that only length is to be minimized, then one might just as well seek trees of minimum Euclidean length, or indeed of minimum length in any of the other uncountably many possible measures of abstract distance. But even that does not exhaust the possibilities. Numerical values—lengths of a sort—can be calculated for branches without regard to the possible features of nodes, by fitting the tree directly to a matrix of pairwise distances between terminal taxa (such methods are reviewed by Farris, 1981). Such trees too might be selected to minimize "length," and this might be done for any of the huge number of ways of arriving at a matrix of pairwise distances. The analogy through length has allowed methods such as this to become confused with parsimony analysis, and that confusion has played a role in specious criticisms of phylogenetic methods. Felsenstein (1981)—one of the main proponents of the idea that "parsimony" might mean almost anything—for example, attributes "parsimony" to one such method that had been used by Edwards and Cavalli-Sforza (1964). Sokal and Rohlf (1981) used a procedure for fitting branch lengths to a distance matrix to analyze the data of Schuh and Polhemus (1980), then criticized the parsimony analysis of the latter authors on the grounds that the distance fitted tree is "shorter." As Schuh and Farris (1981) pointed out, the length that Sokal and Rohlf attribute to their tree is quite meaningless, inasmuch as it is smaller than the number of origins of features required to account for the data (for related discussion see Farris, 1981). That will serve as a general commentary on this class of methods, which are too numerous in their possibilities to discuss here individually. The lengths arrived at by such calculations are generally incapable of any interpretation in terms of origins of features, and the evaluation of trees by such lengths consequently has nothing to do with the phylogenetic parsimony criterion. What is worse, the trees produced by these methods frequently differ in their grouping from parsimonious genealogies, and to that extent the use of these procedures amounts to throwing away explanatory power.

Conceivably, to be sure, there might be some purpose in discarding explanatory power, although it is difficult to imagine a legitimate reason for doing so. But at the least one would want some reason; explanation might be sacrificed in return for something, but, surely, not for nothing at all. None of these interpretations of "length" has been supplied with a rationale of its own. Like many quantitative methods, these seem to have come into being for little reason other than that the possibility occurred to their originators. What justification they have seems to comprise no more than an analogy with parsimony, and that rests only on an equivocation on "length." As that justification is specious, there seems to be no legitimate reason whatever for the use of such procedures in analysis of character information.

Pairwise Homoplasies

The situation is somewhat different with some types of comparative data, such as matrices of immunological distance, in which no characters are directly observed. I have emphasized before (Farris, 1981) that the parsimony criterion cannot be directly applied to such cases, and so I shall not consider them here. (The paper just cited offers other bases for evaluating methods of distance analysis.) Some analogies with distance analytic methods, however, can be related to the present discussion.

In fitting a tree to distances, branch lengths are used to determine a matrix of pairwise tree-derived distances between terminal taxa. The derived distance between a pair is just the sum of the lengths of the branches that he on the path connecting the two taxa on the tree. Evaluation of the fit of the tree to observed distances is based on conformity

of the derived to the observed distance values, this being measured by, say, the sum of the unsigned differences between the corresponding elements of the matrices (other measures are discussed by Farris, 1981). Parsimony analyses can also be used to produce derived distances, the patristic differences of Farris (1967). I had earlier (Farris, 1967) termed the departure of patristic from observed differences (pairwise) homoplasies, and from this, as well as by analogy with distance analytic procedures, one might be tempted to evaluate genealogies according to the total of those pairwise homoplasies (such a suggestion has been made by D. Swofford, pers. comm.). The drawback of doing so is already clear from earlier discussion: the pairwise homoplasies are not independent. Consider a feature common to 20 terminal taxa and a series of putative genealogies that assign those 20 to 1, 2, 4, and 20 separate groups of equal size, so that the number of hypotheses of homoplasy is 0, 1, 3, or 19, respectively. The corresponding sums of pairwise homoplasies are 0, 100, 150, and 190. It is seen that the relationship between independent hypotheses of homoplasy and pairwise sums is strongly non-linear for a single feature. When several features are considered, the total of independent hypotheses is of course found just by summing required homoplasies over independent features, and likewise it develops that the pairwise sum for several features is the total of the pairwise sums for the features taken separately. As the sum of nonlinear functions of variables need not be monotone on the sum of the variables themselves, choosing a tree to minimize total pairwise homoplasy need not lead to minimizing requirements for independent ad hoc hypotheses of homoplasy. Much as for the various length measures just discussed, minimizing pairwise homoplasy seems to have no justification other than that homoplasy ought to be minimized, that is, other than by analogy with parsimony. As this method leads to a departure from parsimony, however, the analogy provides no legitimate justification, and the procedure thus seems to have no logical basis.

A more extreme problem of interpretation of pairwise homoplasies arises in some methods for analyzing electrophoretic data. Suppose that each of three terminal taxa, A, B, C, is fixed for a different allele at some locus, and that these taxa are related through an unresolved tree with common ancestor X. There are a number of ways of calculating distance between gene frequency distributions (see Farris, 1981). To fix ideas, suppose that the Manhattan distance on frequencies is used. The distance between any two terminals is then 2. The ancestor X might plausibly be assigned frequency 1/3 for each observed allele. In that case the distance between X and any terminal is 4/3, the patristic difference between any two terminals is 8/3, and the corresponding pairwise homoplasy is 2/3. That implies that there is homoplasious similarity between any two terminals, but the conclusion is nonsense, inasmuch as there is no similarity between them at all. The three terminals simply have three entirely different conditions of the locus.

The details of that example depend on how gene frequencies are assigned to X, but no assignment can bring all the pairwise "homoplasies" to 0 simultaneously. In part this observation reflects the difficulties inherent in any attempt to utilize distances between gene frequency distributions as evidence in phylogenetic analysis (discussed in further detail by Farris, 1981). Of greater interest for present purposes is what the example reveals about alleles as characters. The algebraic reason for the existence of those spurious homoplasies is that the distance coefficient treats shared 0 frequencies as points of similarity. Two taxa are assessed as similar in that both lack some allele, whereas in fact they simply possess different alleles. It is clear that those shared absences offer no independent assessment of the resemblances among the taxa, as the 0 frequency in any one allele is a necessary consequence of the fixation of any other. This problem, then, results from treating dependent quantities as if they were independent.

That difficulty is not limited to analysis of frequency data. Mickevich and Johnson

(1976) introduced a method in which frequencies are transformed into a two-state coding: any frequency above a cutting point is coded as 1 (presence), any other as 0 (absence). The standard Wagner method is then used to find a tree minimizing required origins of states for the coded data. This procedure obviates many of the difficulties of analyzing frequency data through distance measures, but it still suffers from dependence of variables. Fixation of one allele will necessarily control the codes of others at the same locus. The number of state origins for the coding thus need not indicate the number of independent hypotheses of homoplasy for a genealogy, and this procedure should not then be regarded as a parsimony method. The problem of interdependence can, however, be avoided by choosing a better means of coding. Mickevich and Mitter (1981, 1982) have made impressive progress in developing coding methods for analyzing electrophoretic data.

Covering Assumptions

Inasmuch as the aim is to minimize ad hoc hypotheses, it might seem that one could do better still by posing single hypotheses to cover several separate cases of homoplasy. Any putative genealogy might on that reasoning be defended against any character by concocting some premise to imply that all similarities in that character are homoplasies—or against any set of characters by dismissing evidence in general. I shall refer to such mass dismissals of evidence as covering assumptions. The danger of using covering assumptions can be readily seen through a consideration of usual scientific practice. Suppose that an experiment is designed to evaluate a theory on the basis of readings from several instruments, and that some of the readings do not conform to the theory. If the non-conforming observations are only a few of the many readings made, the theory may seem to offer a generally satisfactory explanation; it is less so to the degree that such observations are abundant. Even then attempts may be made to salvage the theory. If the offending readings all come from the same instrument, and so are logically related, they might be dismissed through the premise that the instrument is defective. (If it is found to be defective, so much the better.) But if no connection can be found among the non-conforming readings, the claim that they are coincidentally erroneous would have to be viewed with suspicion. Even the best theories seldom conform to every relevant observation, and so theories are well founded to the degree that non-conforming observations are rare. If contradictory observations could be dismissed as uninformative without regard to their abundance, the link between theory and observation would be tenuous at best.

Of course this is generally recognized, and so attempts to defend theories by doing away with entire masses of evidence are typically rationalized by postulating mechanisms to account for what would otherwise be coincidental departures of observation from expectation. The legitimacy of that procedure depends crucially on the validity of the postulates used. If the postulated mechanisms can themselves be corroborated by other sources of evidence, their use to defend the original theory is justified, and indeed they constitute improvements or extensions of the original theory. But if such mechanisms cannot be defended on extrinsic grounds, then they amount to no more than ad hoc excuses for the failure of the theory. Logically (albeit not rhetorically) they have no more force than the flat assertion that all non-conforming observations must be erroneous because the theory is true. Covering assumptions must be forbidden in scientific study, not only because they are ad hoc, but more particularly because they provide false license to dismiss any amount of evidence whatever.

The reason for prohibiting covering assumptions might be encapsulated by the observation that their use would allow theories to be chosen without regard to explanatory power. This effect can be seen directly in phylogenetic application. If 20 terminals share

a particular feature, a genealogy consistent with a single origin of that feature explains those similarities fully. A hypothesis of kinship that broke those terminals into two separate groups of 10 would not explain all the similarities among taxa, but it would still explain similarities within those groups. A tree that divided the same terminals into 4 separate groups of 5 would explain still less of the observed similarities, but would still retain some explanatory power, while a scheme that required 20 separate origins would leave the observed similarities entirely unexplained. Some ad hoc rationale might be used to combine 3—or 19—logically independent hypotheses of homoplasy into a "single" hypothesis. The possibility of that combination might be interpreted to mean that all these genealogies but the first conform equally well with observation. If such a course were followed, the differences in explanatory power among the last three hypotheses of kinship would play no role in choosing among them.

Almost any method that led to departure from parsimony might be suspected of involving a covering assumption. One might presume that the various length measures discussed above arise from some underlying premises that would amount to assumptions about the nature of evolution. But inasmuch as those premises, supposing that they exist, have never been made explicit, there is no real possibility of evaluating them as theories, and it is more immediately useful to view those methods as resulting simply from misunderstanding of the explanatory relationship between genealogies and characters. The notion of covering assumptions does, however, offer a way of investigating such methods further. One might attempt to identify conditions under which a certain method would provide accurate inferences on kinship. If the prevalence of those conditions could be corroborated, then this would provide a justification of the method. If no corroboration proved possible, the dependence of the method on a covering assumption would at least have been established.

Felsenstein's maximum likelihood methods offer fine examples of reliance on covering assumptions. The stochastic models would—if they were realistic—explain why seemingly independent characters would depart systematically from a parsimonious arrangement, and so would justify preference for unparsimonious schemes. That neither Felsenstein nor anyone else maintains that those models can be corroborated, likewise, makes it clear that in practice that justification would be entirely specious. But most of these methods have never been advocated for practical application, anyway. Felsenstein's own recent efforts center on likelihood interpretations of procedures that had already been advocated on other grounds, as I shall discuss later. It is of more practical interest to analyze methods that have been proposed more or less seriously.

Irreversibility

Some techniques have been proposed as restricted "parsimony" methods. In these the number of origins of features is minimized, but subject to the condition that some kinds of origins be rare or forbidden. Commonly, methods of this sort embody some version of the idea that evolution of individual characters is irreversible. In the method of Camin and Sokal (1965) secondary plesiomorphies are supposed not to occur, and so are excluded from reconstruction, the tree being chosen to minimize parallelisms. In the "Dollo" method of Farris (1977) origins of apomorphic traits are supposed to be unique—traits once lost cannot be regained—and the tree is chosen to minimize secondary plesiomorphies. The Camin-Sokal assumption also crops up in other contexts. As Farris and Kluge (1979) demonstrate, some clique methods tacitly utilize that premise. Monothetic group methods (reviewed by Farris, Kluge, and Mickevich, 1982) do likewise. I shall discuss the former separately. In the latter, putative monophyletic groups are required to be monothetic in the sense of Sharrock and Felsenstein (1975), which is to say that they must possess combinations of features that are distinctive in a certain,

restricted sense. Farris, Kluge, and Mickevich show that the monothetic requirement leads (in present terms) to loss of explanatory power. As they put it, the method imposes otherwise unnecessary requirements that observed similarities be dismissed as coincidental. No one has suggested any rationale for the supposition that monophyletic groups must be monothetic; the method seems to be based on an esthetic principle, if anything. It thus seems to merit no further discussion.

Any of these methods might yield the same genealogy as would be obtained without the restriction, but none of them needs do so; and in general applying the restriction will increase the number of hypotheses of homoplasy needed to defend the conclusion. There is no particular limit to that increase, so that using the restriction amounts simply to dismissing en masse any evidence that might otherwise seem to vitiate the conclusion.

It is of course entirely conceivable that unidirectional evolution is true of some characters, but the possibility is hardly enough to justify irreversibility as a prior restriction on conclusions. The restriction has been applied, moreover, in analyses of characters, such as morphometric traits, for which a guarantee of directed evolution seems quite implausible. (Blackith and Blackith, 1968, did so, then dismissed phylogenetic analysis in general on grounds of the obvious absurdity of their own premises.) The motivation for doing so seems to be a matter more of technical convenience than of conviction of the propriety of the restriction. That seems particularly to apply to the Camin-Sokal method, as it was one of the earliest techniques to be implemented as a computer algorithm. The reason for my own (Farris, 1977) development of the Dollo method likewise had little to do with the realism of the assumption. That study was intended primarily to show logical flaws in Le Quesne's (1969, 1974) earlier attempts to analyze the same problem.

In a serious study, defending conclusions that depended crucially on use of a restricted method would require defending the restriction itself. I would not claim that the supposition of irreversible character evolution could never be supported by extrinsic evidence. I would suggest, however, that what acceptance that idea has gained has been based mostly on generalizations derived from hypotheses of kinship. The common notion that evolution generally proceeds from many, similar, parts to fewer, differentiated, parts, for example, seems to have been arrived at by induction from putative lineages. If the putative phytogenies used to draw such conclusions had been arrived at by presupposing irreversibility, then the conclusion would have no legitimate empirical support. If the idea of irreversibility is supported at all, then, it must have been derived from analyses that did not depend crucially on its truth. The evidence for a directed evolutionary trend, then, would be that the postulated trend conforms to a pattern of kinship that is in turn supported by other evidence, that is, that itself conforms to other characters.

Naturally it could be charged that, by allowing reversals, parsimony effectively makes the assumption that reversal is possible—that irreversibility is not true. That is so in a sense. If it were known that evolution is irreversible, application of that knowledge might lead to genealogical inferences that otherwise might seem unparsimonious. But in fact no such thing is known, and the attempt to apply an empirically supported claim of irreversibility as a criticism of parsimony leads to a peculiar difficulty. Any body of characters might be made to appear to be consistent with the postulate of irreversibility. It is always technically possible to construct a tree so that all homoplasies take the form of parallelisms. It might seem from this that character information could never challenge the theory. But if the evidence for irreversibility was originally based on character distributions, then it would be quite unwarranted to analyze further cases so as to force them into conformity with irreversibility. The effect of doing so would be precisely to confer on irreversibility the status of an empirical conclusion that cannot be questioned by evidence—a contradiction in terms. In order to avoid that fallacy, it is necessary to

allow that character information may support a conclusion of reversal. Whenever a putative reversal offers a more complete (that is, as already seen, more parsimonious) explanation of observed similarities than does a reconstruction enforcing irreversibility, irreversibility must be discounted (in that particular case; trends might still be accepted as rough descriptive generalizations).

A proponent of irreversibility might nevertheless insist that when an analysis that does not presuppose irreversibility gives a different result from another that does use that premise, then the conclusion of the former depends crucially on the supposition that reversal is possible. As the procedure just outlined, moreover, will always discount irreversibility when parsimony requires, there is no way of rejecting the possibility of reversal. That possibility might seem, then, to be an ad hoc hypothesis, so that a conclusion of reversal actually requires more ad hoc hypotheses than would be suggested just by counting independent origins. But even if possibility of reversal did constitute an ad hoc hypothesis, it would certainly not be an additional independent hypothesis, for it is entailed by the particular hypothesis of reversal postulated. That observation, in fact, contains the key to the defect of the whole objection. If a particular conclusion of reversal could be legitimately criticized as presupposing the possibility of reversal, then any scientific conclusion whatever could be dismissed as requiring the supposition of its own possibility. The argument outlined is seen in that light to be simply another rationalization for discarding evidence.

It will be worthwhile to pursue that conclusion in further detail. It is clear that the reasoning outlined effectively views irreversibility and the possibility of reversal as competing theories. The charge that possibility of reversal cannot be rejected by parsimony analysis would be pertinent only as a criticism of a way of testing an empirical claim. But that view is itself suspicious. Irreversibility is certainly an empirical claim, and further it is plainly testable in principle, inasmuch as it prohibits something, namely reversals. Possibility of reversal, on the other hand, can hardly be by itself an empirical claim in the same sense (although the claim that particular reversals have occurred might be), as it does not prohibit anything. One might think that admitting that reversal might occur, if not itself directly an empirical contention, nonetheless implies one, in that using a method that can discard irreversibility for parsimony would necessarily yield conclusions of reversal. But in fact it is quite possible for a parsimonious reconstruction to lack requirements for reversal. (The contrary, of course, is also possible, and is often observed. But that is a consequence of the idea in conjunction with particular observations, not of the idea itself.) While irreversibility and the possibility of reversal seem superficially to be simply alternative theories, then, they are in fact not the same kind of idea. The first is a theory that forbids conclusions that might otherwise seem supported by observation, and, when confronted with such cases, can be saved only by ad hoc supposition. The second is simply an attitude. The possibility that irreversibility (or any theory) is false must be considered in order to test the theory. No kind of empirical science would be possible without such attitudes.

Because of their reliance on covering assumptions to justify otherwise unnecessary ad hoc hypotheses of homoplasy, the Camin-Sokal and Dollo techniques should not be regarded as proper parsimony methods, prior usage notwithstanding. The situation may be different, however, for another restricted procedure. In the chromosome inversion model of Farris (1978) it is presumed that each of two alternative inversion types originated uniquely. Inversion types may nonetheless show incongruence with a genealogy through independent fixations from polymorphic ancestral populations, and the tree is chosen to minimize such fixations. The accuracy of the premise of unique origin might of course be questioned, but the idea is accepted by specialists on grounds extrinsic to genealogical hypotheses, and I shall not attempt to argue its validity here. A further observation in this connection, however, seems worthwhile. As this model presumes a

unique origin for each inversion type, it might seem that similarity between organisms would on this premise be due to inheritance regardless of the genealogy postulated, so that the relationship between parsimony and genealogical explanation would no longer hold. The inherited similarity covered by the premise, however, holds only for chromosomes of individuals. Resemblance between populations fixed for the same inversion may still be explained by inheritance, or else the coincidental result of independent fixations. As it is populations that are grouped in postulating a genealogy, it is still possible to compare alternative genealogical hypotheses on explanatory power. There is in fact nothing unusual in this conclusion. It is generally true that features used to arrange taxa are characteristics of populations, rather than of individuals. The observation that deer have antlers is just a contracted way of stating that normal adult male deer in breeding condition possess those structures. The females, young, and deformed are not given a separate place in the system by reason of lacking the characteristic. The same principle underlies Hennig's emphasis of the idea that holomorphs, rather than specimens, are classified. Mickevich and Mitter (1981) arrive at the same concept in developing their greatly improved methods for analyzing electrophoretic data. They concentrate on recognizing suites of alleles as features of populations, rather than attempting to use single alleles—traits of individuals—as characters.

Phenetic Clustering

Clustering by raw similarity (phenetic clustering) has sometimes been advocated as a means of making genealogical inferences, typically with the justifying assumption that rates of evolutionary change (or divergence) are nearly enough constant so that degree of raw similarity reflects recency of common ancestry. The method is most often used with comparative biochemical data, but it has also sometimes been recommended for morphological data as well (for example, by Colless, 1970).

Constancy of rate is rather a different theory from irreversibility of evolution, but many of the comments made earlier apply here as well. Phenetic clustering might coincidentally produce a parsimonious scheme, but it certainly need not do so, and again there is no limit in principle to the number of otherwise unnecessary requirements for hypotheses of homoplasy that this method might impose. The assumption is certainly an empirical claim, and advocates of the method usually defend it by producing evidence for rate constancy. (Colless is an exception; he shows no inclination to resort to evidence). That evidence typically takes the form of correlations between observed raw similarities and putative recency of common ancestry. Those last naturally depend on hypotheses of kinship, and this raises the familiar dilemma. If the genealogies used as evidence depended crucially on rate constancy, there would be no evidence. Supposing, then, that they do not, the evidence must consist of agreement between the theory and arrangements that conform to character distributions. Just as before, if the premise of rate constancy is used to justify unparsimonious conclusions, the effect is to consider rate constancy as empirical and irrefutable at once. Likewise parsimony analysis might be accused of presupposing that rates can vary, but discussion of that idea would precisely parallel what has already been said in connection with irreversibility.

The rate constancy idea nonetheless has some defects all its own. Character data may be converted into degrees of raw similarity in any of a great number of ways. If some of those measures exhibit rate constancy, it is too much to expect that they will all do so. The claim that raw similarity is well correlated with recency of common ancestry has no clear meaning unless the measure of similarity is specified. Defense of the theory then rests on justification of a similarity measure. Particularly for molecular data, both are often defended at once through stochastic models of evolution. Nei (1972) derived his measure of genetic distance for electrophoretic data from the premise that evolution

comprises a stochastically homogeneous process of (eventual) fixation of (nearly) neutral alleles.

The obvious comment is that such derivations no more justify anything than does one of Felsenstein's exercises, unless it is supposed that the model is realistic. In this case, however, the situation is even worse. As a stochastic model is invoked, one might well demand that the clustering method be statistically justified, not just the similarity measure. Cavalli-Sforza and Edwards (1967) used a genetic drift model, too, and found that clustering by raw similarity need not yield the tree optimal under the model. The reason for this is that even when evolution is stochastically homogeneous, observed similarities between particular populations need not closely reflect recency of divergence. Stochastic variation is to be expected in realized degrees, of evolutionary change, though the average rate is uniform. Two populations less similar to each other than to a third might then nonetheless form a genealogical group excluding the third population. The model intended to justify phenetic clustering does not do so, even if its premises are accepted.

Phenetic clustering might still be salvaged by adding still further justifying assumptions. If speciation events were widely enough separated in time, it would be unlikely that stochastic fluctuations in a homogeneous change process would be large enough to produce a genealogically false clustering. But if long periods of time are invoked, the model itself might be questioned simply on that basis. It might well be that evolution is nearly a homogeneous process over short time intervals, yet still shows sporadic changes in rate during longer periods. If such possibilities are taken into account, it is seen that phenetic clustering rests on a whole series of suppositions. It seems pointless to pursue such premises in further detail. They might be summarized as the claim that evolution operates so that genealogically most closely related taxa are also mutually most similar.

That claim can be evaluated directly, without references to the details of models. That is just as well, as the premises of most such models tend to the incredible, anyway. Most of the evidence used to support the theory of homogeneity of rate, in fact, is not actually such, but instead more directly tests the more general claim. A correlation between similarity and putative recency of common ancestry, for example, might be interpreted as showing rate homogeneity, but it is also consistent with the possibility that time intervals between speciations have combined to cancel the effects of rate fluctuations. There are other possibilities still, but for present purposes there is no need to distinguish among them. For evaluating clustering by raw similarity, it is just the relationship between kinship and raw similarity that is of interest.

A molecular evolutionist is quite happy with the generally good correlation that is observed between raw similarity and putative recency of common ancestry; for him it substantiates the molecular "clock." But, as I have emphasized before (Farris, 1981), such correlations are not enough to justify clustering by raw similarity. The correlations reported show considerable scatter. The implication of this, accepting the usual interpretation of the general correlation, is that rates of divergence vary somewhat. Even if it is generally true, then, that genealogically most closely related taxa are also mutually most similar, there are evidently exceptions. These exceptions could not be identified if genealogy were inferred by presupposing that raw similarity reflects kinship. To make accurate inferences in such cases—to discover what the cases are—it is necessary to use a method that can discount raw similarity as indicative of kinship if the data seem to require doing so—if doing so is required to achieve a more complete explanation of observed features. By analogy with the discussion of irreversibility, the same conclusion would be reached just by requiring that the relationship of raw similarity to kinship be vulnerable to evidence. It seems, then, that the theory that raw similarity indicates kinship—even if often accurate—can provide no legitimate grounds for accepting unparsimonious inferences.

I commented before on the distinction between an ad hoc covering assumption and a

corroborated improved theory able to account systematically for observations that would otherwise seem coincidental departures from its predecessor. This distinction suggests a further defect in the attempt to defend phenetic clustering on grounds of a correlation between raw similarity and kinship. In a legitimate extension of theory, the old coincidences are not dismissed as such, but explained by the extension. The process, that is, expands explanatory power, rather than discarding it. Suppose that clustering by raw similarity in some case requires otherwise unnecessary hypotheses of hompolasy, and that the conclusion is defended on grounds of a theoretical relationship between raw similarity and recency of common ancestry. If this is not ad hoc, then the theory must offer an explanation of the putative homoplasies. It is far from clear, however, that it can do so. Homoplasies, as already observed, are not explained by the inferred genealogy, from the standpoint of which the shared features that they represent are so only coincidentally. Inasmuch as raw similarities are calculated from features, it seems curious that they could either explain or be explained by a scheme that left the features themselves unaccounted for. In order for a relationship between raw similarity and kinship to explain homoplasies, it would seem necessary to suppose that the relationship rests on some real mechanism. That mechanism would have to have the property that organisms would come to possess features in common for reasons other than inheritance, and in just such a way as to maintain the correlation between raw similarity and recency of common ancestry. As no known natural process appears to have this property, it would seem that use of a postulated correlation between raw similarity and kinship to defend clustering by raw similarity rests necessarily on an ad hoc covering assumption.

A related conclusion can be reached by another route. Phenetic clustering ignores considerations of parsimony and so effectively proceeds by freely introducing whatever hypotheses of hompolasy are needed to derive a result conforming to the rate constancy premise. The procedure would be highly questionable on statistical grounds alone, then, if homoplasy were supposed to be rare. The method then requires the assumption that homoplasy is abundant, and indeed its proponents are prominent in criticizing parsimony as requiring rarity of homoplasy. The premise that homoplasy is abundant, however, poses a problem for clustering by raw similarity as well. That method infers recency of common ancestry of two taxa from the fraction of characters in which the two are similar. If homoplasy were rare (and rates constant), that would be superficially reasonable. Similarity between two lineages would decrease in clocklike fashion as ancestral similarities were lost. But if homoplasy is abundant, many of the similarities between two taxa are likely to be homoplasious, in which case they need indicate nothing about how recently the pair diverged. Two populations having only a remote common origin, and so (if rates were constant) very little homologous similarity, might have many recently acquired homoplasies, and so be judged to be of recent common ancestry. It is easy enough to identify conditions under which inferences might still be valid. If pairwise homoplasies were all the same, or nearly so, homoplasy would not alter the relative degrees of raw similarity among taxa, and then (if rates of change were constant) the method would still work. Phenetic clustering effectively presumes, then, that the variance of pairwise homoplasies is small. Keeping that variance small would be the task of the hypothetical mechanism just discussed.

While phenetic clustering does not consider homoplasies as such, it does select a tree by finding a constant rate (ultrametric) model that conforms to observed raw similarities as closely as possible. If rates of evolution were constant, homologous similarities would conform to the constant rate model, so that departure from the model would be due to variation in pairwise homoplasies. The phenetic clustering procedure most commonly applied for genealogical inference, UPGMA, has precisely the effect of minimizing the variance of pairwise departures of observed from ultrametric similarities (Farris, 1969a).

Phenetic clustering and parsimony analysis are similar, then, in the sense that each minimizes a criterion. But whereas abundance of homoplasy need not imply error by parsimonious inference, large variations in pairwise homoplasies would certainly vitiate the conclusion of phenetic clustering. Phenetic clustering, unlike parisimony, depends crucially on minimality in nature of the quantity that it minimizes. Clustering by raw similarity possesses the very defect that its proponents had incorrectly claimed as a weakness of phylogenetic analysis.

Cliques

Clique methods rely on parsimony to interpret suites of congruent characters, but their trees require homoplasy for characters outside the selected clique, and often the clique tree will be quite unparsimonious for those characters. In practice the excluded characters are often numerous, so that basing the inferred genealogy just on the clique imposes a considerable loss of explanatory power. These methods are then prime suspects for reliance on a covering assumption, but for a long time it was not clear from the clique literature what that assumption was supposed to be.

Advocates of clique methods have put forward a variety of ostensible reasons for adopting that approach, but consideration of those discussions leads quickly to the suspicion that they were intended primarily to obscure the underlying premise of the procedure. A number of authors (reviewed by Farris and Kluge, 1979) have claimed that clique methods are Hennigian, whereas parsimony analysis is not. But that contention was never documented by citations from Hennig, and, as Farris and Kluge pointed out, that idea is directly contradicted by Hennig's insistence that the phylogenetic system be brought into conformity with all available evidence. In the more recent literature (Duncan et al., 1980) a connection between Hennig and cliques is no longer claimed. (Estabrook, cited by Funk and Brooks, 1981, has gone a step further than that, now taking the position that cliques are not intended for phylogenetic analysis.) Together with the earlier lack of documentation of the Hennigian basis of cliques, this suggests that the claim was never more than a subterfuge. Similar inconsistencies are found in other rationalizations. Duncan et al. contend that clique and parsimony methods usually give similar results. This makes for an interesting contrast with the emphasis by Estabrook et al. (1977) of differences in results of the two procedures. The claim is of course false, but even if true it would obviously provide no reason for departing from parsimony. Other arguments go further along the path of defending cliques by pretending that they amount to parsimony. Some authors (again see Farris and Kluge) have portrayed that method as uniquely capable of detecting homoplasy or of providing a means of analyzing characters. Those abilities are of course characteristics of parsimony analysis. The contention that clique methods allow detection of homoplasy, furthermore, is at best weak. Supposing their inferences to be well founded, those methods might specify that excluded characters showed homoplasy, but that would scarcely indicate which similarities in those characters were homoplasies, or which were not. Duncan et al. give an example in which, they say, clique analysis improves on parsimony by revealing two alternative interpretations of conflicting characters. But the two are in fact equally parsimonious genealogies. Estabrook and Anderson (1978), likewise, recommended clique methods on the basis of highly satisfactory results, but, as Farris and Kluge document, those results had in fact been obtained by parsimony analysis.

Le Quesne (1972) offered an approximate method (later made exact by Meacham, 1981) for finding the probability that a suite of characters would all be congruent if features were distributed independently and at random among taxa. He suggested selecting the clique with the lowest such probability, and other proponents also commonly refer to cliques as "least likely." It is possible that this idea is intended as a justification

of clique methods. If so, the justifying reasoning amounts to no more than misunderstanding of statistics. If a clique were evaluated just on its probability under a random model, the evaluation would be bound to the model. In that case the covering assumption of cliques would be that characters—being randomly distributed—have no relationship to genealogy. Perhaps it was intended that low probability under a null model would lend credence to an alternative, genealogical interpretation of the clique, but that idea, too, rests on a fallacy. Observing that a clique (or anything else) has low probability under a model might provide statistical grounds for rejecting the model, but it does not by itself offer any basis for choosing an alternative hypothesis. Once a model has been rejected, the probabilities it assigns to events necessarily become irrelevant. In this case rejecting the null model is uninformative, as no one interested in making phylogenetic inferences would have taken it seriously anyway. The statistical reason for accepting a new hypothesis is that it assigns much higher probability to observation than does the old. In normal statistics, a large enough difference between sample means serves as grounds for rejecting the hypothesis that the two samples were drawn from populations with the same parametric mean. If an alternative hypothesis is chosen so that it assigns maximum probability to the observed difference, the new theory conforms best with observation. But one hardly proceeds by choosing observations so as to minimize their nominal probability under the original hypothesis, let alone using such observations as the basis for choosing a new theory. Making statistical genealogical inferences from characters that had been used to reject the hypothesis of randomness would likewise require choosing a genealogy that would assign maximum probability to available characters—a maximum likelihood tree. I have already commented on the difficulties of applying that approach in practice, but this case is far worse. No model other than the rejected one of randomness is provided, and so neither are any grounds whatever for accepting the tree from the "least likely" clique as a genealogical inference best conforming to observation. (Felsenstein has made much the same point.)

As none of the ideas just discussed provides any legitimate rationale for clique methods, those procedures must rest on an undisclosed assumption, if indeed they rest on anything at all. It is not difficult to discern what that assumption would have to be. Cliques are usually chosen to comprise as many mutually congruent characters as possible and any characters that must be discarded to achieve this are simply counted as excluded. If the genealogy corresponding (by parsimony) to the clique is accepted, each of the excluded characters will require at least one hypothesis of homoplasy, but the number required may well vary among those characters. As characters are counted just as excluded or not, the number of hypotheses of homoplasy required by excluded characters plays no role in the analysis: similarities in those characters are dismissed en masse. The covering assumption involved is thus like the archetypical one discussed before. Ad hoc hypotheses of a sort are counted, but the counts do not reflect simple hypotheses of homoplasy. Instead any and all similarities in each excluded character are discounted by recourse to a "single" covering assumption. Excluding a character amounts to treating all similarities in it as irrelevant to assessing kinship. Those similarities could all be logically irrelevant only if they were all homoplasies. The covering assumption utilized is then that excluding a character—concluding that it shows some homoplasy—implies that all points of similarity in that character are homoplasies.

As discussed before, the collective dismissal of similarities in a character would be justified if the multiple required origins of features were not logically independent. It is readily seen, however, that such is not the case. The conclusion that endothermy has evolved independently in mammals and in birds does not imply that each species of bird or mammal has independently achieved that condition. Such being the case, it is likewise already clear from earlier discussion that use of such a covering assumption leads to loss of explanatory power. As before, a single requirement of homoplasy may leave

many of the similarities in a character explained, while a large enough number of required homoplasies will leave the same similarities entirely unexplained. Counting characters as simply excluded or not produces an evaluation oblivious to that distinction.

This defect of cliques is not entirely general. In a case with only three terminal taxa, any single apomorphy will require at most a single hypothesis of homoplasy. As classifying characters simply as homoplasious or not then discards no information, the clique method would coincidentally yield the same result as parsimony analysis. The clique tree may well depart from the most explanatory genealogy with larger numbers of terminals, but as a rule the departure is less, the fewer the terminals considered. Some modifications of the clique method take advantage of this phenomenon by forming cliques for each of several subsets of terminals, then fitting the several trees together in some way. To the extent that the use of subsets of taxa diminishes the reliance of the method on its covering assumption, this approach is beneficial, but the validity of the method for fitting parts of the tree together then becomes crucial. Farris and Kluge (1979) showed that some versions of this process rely on the further covering assumption of irreversibility, and such methods would have to be rejected on that score alone, as has already been seen. Neither the procedures for selecting the subsets and combining trees nor any grounds on which either might be justified have been clearly set out, and as Farris and Kluge document, some published examples incorporate contradictory principles. There is, moreover, some reason to suspect that the cutting and fitting is often performed by trial and error, with the unadmitted aim of bringing the result close to that of a parsimony analysis. The lack of explicit formulation of the procedure, in any event, makes it impossible to discuss further.

Advocates of other unparsimonious methods, such as phenetic clustering, attempted to defend their preferences more or less forthrightly, by advancing covering assumptions as respectable theories. The clique proponents cited above did not do likewise, but instead resorted to discussion aimed at obscuring the covering assumption of that method. The nature of that assumption makes it clear, I think, why those advocates were driven to take that course. To attempt to justify clique procedures on their actual basis would have served only to invite ridicule. Unlike at least superficially tenable rationalizations such as rate constancy, the clique assumption implies a claim that no one would seriously entertain as a realistic theory. Lack of realism has never bothered Felsenstein, though. He has proposed two stochastic models (Felsenstein, 1979, 1981) under which he derives cliques as maximum likelihood estimates of genealogy. Both of these operate, just as would be expected from the clique assumption, by supplying principles that would excuse dismissing characters as units. In the 1979 paper, this effect is achieved by introducing the possibility of a carefully selected type of error. Any character incongruent with the accepted tree is characterized as erroneous, and this is taken to mean that the character has been so completely misinterpreted that it is uninformative on genealogy. In the other model, incongruent characters are instead regarded as having changed so frequently in evolution as to be unrelated to genealogical grouping.

Both models incorporate uncorroborated—and quite unrealistic—theories of evolution, and I shall take that point up below; but it is of more immediate interest to discuss the qualities of the two premises that relate the models to cliques in particular. Both rationalizations of the clique assumption, it will be seen, have the same defect as the clique assumption itself.

The error idea rests on a misrepresentation of how systematists recognize characters. It was at one time believed that the eyes of octopi and of vertebrates were the same. That was certainly an error, for the two organs differ in both structure and ontogeny. But that that mistake was made does not mean that the sameness attributed to the eyes of rats and of mice is likewise a misinterpretation. Concluding a homoplasy in a feature may well invite renewed inspection and possible reinterpretation, as already discussed.

But that reinspection certainly need not lead to dismissing all the agreements between taxa that the original feature had been intended to summarize. While it is reasonable to attribute some homoplasies to errors, it does not follow from this that those errors will turn out to have universal effects. It is seen from this that Felsenstein's use of the idea of wholesale error as a defense of the clique assumption amounts to no more than stating the desired conclusion as a premise. The defect of cliques is just that they treat every conclusion of homoplasy as if it implied universal homoplasy. Felsenstein attributes homoplasies to errors, but bolsters cliques only by supposing that any conclusion of error implies universal error. Neither implication is valid, and so either is merely an ad hoc rationalization for dismissing relevant evidence.

Much the same applies to Felsenstein's second argument. Dismissing an incongruent character on the grounds that it must have changed very frequently clearly depends on discounting the possibility that it changed only a few times. As Felsenstein (1981:183) puts it, the clique method is suitable ''when it is known that a few characters have very high rates of change, and the rest very low rates, but it is not known which characters are the ones having high rates." He does not disclose, however, how one comes to know the rate of change of a character without a prior phylogenetic analysis. Nor does he explain how one would apply that undisclosed method to gain the knowledge that his method calls for, without in the process incidentally learning which characters had the high rates. Nor, again, does he offer any pretense of a reason why rates should restrict themselves to be either very high or else very low—or why rapidly changing characters ought to be "few." In the absence of such explanation, it is seen that the covering assumption that one conclusion of homoplasy implies universal homoplasy has once again been "defended" simply by restating it as the entirely equivalent—and equally unsubstantiated—premise that any feature that originates more than once must have done so a very great number of times.

It seems unnecessary to debate that supposition further, as Felsenstein himself emphasizes that the premises of his model are inconsistent with the observed abundance of incongruence among characters. It seems worth noting, however, that there is a conflict between Felsenstein's idea of errors and his second model. Returning to the eyes of octopi and vertebrates, the old interpretation of identity would yield a feature incongruent with many other characters. Under the frequent change interpretation, then, this would be taken to mean that vertebrate eyes had originated many times. But once the identity of the two sorts of eyes is recognized as an error, it is seen that two different features are involved. Then the "conclusion" of many origins of either one vanishes: it was based simply on an error. Of course the relatively simple mistake of misidentifying octopod eyes seemed to lead to such an astounding conclusion only because it was interpreted according to an equally astounding premise. The specious conclusion could be avoided simply by requiring that conclusions of multiple origin—of difference between structures that seem the same—like other scientific conclusions, be supported by evidence. That requirement is Hennig's auxiliary principle, which is to say it is parsimony.

Likelihood

An attempt to defend clique techniques would seem to imply a criticism of other methods, but Felsenstein's 1981 paper does not on first inspection seem to offer such a view. Instead variations of a stochastic model are used to derive a variety of procedures. He arrives at parsimony by supposing that character changes—and so homoplasies—are very infrequent generally, the alternative of rapid change in some characters being neglected for this purpose. Weighting as a function of rates of change that differ among characters is also derived, again with the restriction that the rates, while variable, are

all extremely small. Felsenstein suggests that this result justifies the old idea that characters ought to be weighted according to conservatism—slow rate of evolutionary change. He also uses a variation of his clique derivation to produce new ("threshold") methods. These differ from cliques just in that a character may have any number of homoplasies up to a specified limit (the threshold) before it is dismissed as rapidly evolving (cliques are then threshold methods with threshold 1). These of course have the same defect as cliques, if perhaps in lesser degree: beyond some point, further loss of explanatory power is neglected, and the dismissal of evidence is rationalized by recourse to an unsupported assumption. In this case Felsenstein himself points out that the assumption is not only unsupported, but refuted by available evidence. He nonetheless recommends his new methods as "of use in analyzing real data."

Felsenstein relies on maximum likelihood for all his derivations, and he characterizes his favorite new method in particular as having the advantage "that it springs organically from a likelihood approach." One would think from that last that maximum likelihood methods have distinctive advantages, and from all this, that maximum likelihood provides a justification for all the methods Felsenstein derives. But that impression conflicts sharply with his earlier arguments. In 1978 (p. 408) he emphasized, "Maximum likelihood estimates are not desirable in themselves, but because they have desirable properties, such as consistency." In the same paper, as already has been discussed, he argued that both clique and parsimony methods are inconsistent—and that development would apply to threshold methods as well. Choice among the various methods that Felsenstein now derives from maximum likelihood, furthermore, requires, according to his derivations, knowledge of such evolutionary parameters as what fraction of characters change rapidly. Using any of these methods, moreover, requires knowing that no characters change at moderate rates. But in 1979 (pp. 60-61) Felsenstein had stressed that maximum likelihood methods cannot be used in practice because "that would require knowledge of the parameters of the model." At the best, finally, one would have to say that all these derivations rest on rather questionable assumptions. But the 1978 paper concludes, "If phylogenetic inference is to be a science, we must consider its methods guilty until proven innocent."

One might well conclude from all this that Felsenstein has realized that his earlier arguments are nonsense. But he offers no such comment, and I suggest that instead he has fallen into these contradictions accidentally, in the course of pursuing a new aim. That goal is to establish the idea that the various methods he discusses all rest necessarily on the same basis. In order to accomplish this, he ignores or attempts to gloss over any differences among the bases of those procedures. He portrays the relationship between weighting and frequency of change (or conservatism) as a new conclusion, one derived just from his models. He does not mention that Farris (1966, 1969b, 1977, 1978) and Kluge and Farris (1969) had already made similar suggestions on other grounds. Having argued in 1978 that both cliques and parsimony methods are inconsistent, he cites that paper later (1979:61), but with a new interpretation: both methods "seem to have passed the few tests applied so far." In 1981 he goes further toward concealing the distinction between cliques and parsimony: the former "can be viewed as parsimony methods, in that one is counting number of characters incompatible with the tree" (p. 191). Of the relationship between origins of features and hypotheses of homoplasy he writes, "It would be easy to apply the same argument to [cliques] and count number of characters incompatible. . . . Given that we are in doubt as to how to quantitate the number of hypotheses, it seems to be best to rely instead on a more traditional framework, that of statistical inference." In applying the astounding views that there is no difference between using covering assumptions and not doing so and that statistics is a more traditional basis of empirical investigation than is conformity to evidence, Felsenstein arrives at both clique and parsimony methods in just the same

way. He derives both from premises that are—as he himself emphasizes—unrealistic. The effect of his discussion is nothing other than to create the impression that clique and parsimony methods are equally "well" justified, that is, that parsimony is not justified at all.

That position is nothing but pretense, for parsimony possesses justifications that clique methods lack. My own statistical arguments lead to minimizing origins, not incongruent characters. Felsenstein perhaps feels that those derivations can be ignored because of his objections to them. But his objections concern only consistency, and the faults of that line of reasoning have already been seen. As his own new method has the same "defect," at any rate, his continued recourse to that argument would embody inconsistency in a more common sense. But while my derivations serve to show the logical weakness of Felsenstein's position, they have little further significance, as I have already explained. Far more important is the relationship of parsimony—but not cliques or thresholds—to ad hoc hypotheses, and so to explanatory power. Felsenstein tries to obscure this, too, but his "doubt" on how to count ad hoc hypotheses is based on no argument; it is merely a confession—or pretense—of ignorance. It serves only to supply a facade of rationality for the stance that scientific considerations may be ignored, and that ludicrous premises may be used to "justify" methods of phylogenetic inference, because the statistician can manage nothing better.

Conclusion

Clique advocates are not alone in portraying parsimony as depending on unrealistic assumptions. Advocacy of all other non-phylogenetic methods has consistently been based on the same charge, as has opposition to phylogenetic systematics generally. That allegation has never been supported by substantial argument. It has been motivated instead by the dependency of other approaches on false suppositions: proponents of other views have tried to bolster their positions through the pretense that no means of phylogenetic analysis can be realistic.

Parsimony analysis is realistic, but not because it makes just the right suppositions on the course of evolution. Rather, it consists exactly of avoiding uncorroborated suppositions whenever possible. To a devotee of supposition, to be sure, parsimony seems to presume very much indeed: that evolution is not irreversible, that rates of evolution are not constant, that all characters do not evolve according to identical stochastic processes, that one conclusion of homoplasy does not imply others. But parsimony does not suppose in advance that those possibilities are false—only that they are not already established. The use of parsimony depends just on the view that the truth of those—and any other—theories of evolution is an open question, subject to empirical investigation.

The dichotomy between parsimony and supposition is just that; parsimony offers no barrier to evolutionary theories as such. Rate constancy—or any other supposition-seems to be in conflict with parsimony in the abstract, as it seems to offer a different basis for making genealogical inferences. But it would conflict with parsimony in application only in conjunction with observation, if maintaining the supposition required discarding a parsimonious interpretation of evidence. In that case, however, the same evidence would serve to question the supposition, which could then be defended only by presupposing its truth, or—entirely equivalently—simply dismissing the evidence. But if parsimonious interpretation of evidence did not refute the supposition, then the latter would become a corroborated theory. Parsimony does not require that no such theories will be corroborated, but offers a means for that corroboration, provided evidence allows it. Unlike prior supposition, empirically supported evolutionary theories can offer no criticism of parsimony, for those theories could have become corroborated just to the extent that they require few dismissals of evidence. The insistence by pro-

ponents of suppositions that parsimony is unrealistic, it is then seen, is merely a subterfuge. Ostensibly the objection is to parsimony, but in fact the complaint is that some cherished idea does not conform to evidence.

I return finally to the questions raised at the beginning. Phylogenetic analysis is necessarily based on parsimony, both because it is precisely that criterion that leads to grouping according to putative synapomorphy and because empirical investigation is impossible without avoiding ad hoc hypotheses. Only synapomorphy provides evidence of kinship, for the attempt to use raw similarity as evidence necessarily either would rest on uncorroborated—and so non-evidential—supposition or else could lead to no conclusion conflicting with synapomorphy. And phylogenetic analysis is most certainly empirical, for in applying the parsimony criterion,. it chooses among alternative hypotheses of relationship on nothing other than their explanatory power. Differing as it thus does from all other approaches, phylogenetic systematics alone provides a logical basis for the empirical study of the relationships among organisms.