Philosophical Conjectures and Their Refutation

ARNOLD G. KLUGE

Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109-1079, USA;
E-mail: akluge@umich.edu

Abstract.—Sir Karl Popper is well known for explicating science in falsificationist terms, for which his degree of corroboration formalism, $C(h,e,b)$, has become little more than a symbol. For example, de Queiroz and Poe in this issue argue that $C(h,e,b)$ reduces to a *single* relative (conditional) probability, $p(e,hb)$, the likelihood of evidence e, given both hypothesis h and background knowledge b, and in reaching that conclusion, without stating or expressing it, they render Popper a verificationist. The contradiction they impose is easily explained—de Queiroz and Poe fail to take account of the fact that Popper derived $C(h,e,b)$ from absolute (logical) probability and severity of test, $S(e,h,b)$, where *critical evidence*, $p(e,b)$, is fundamental. Thus, de Queiroz and Poe's conjecture that $p(e,hb) = C(h,e,b)$ is refuted. Falsificationism, not verificationism, remains a fair description of the parsimony method of inference used in phylogenetic systematics, not withstanding de Queiroz and Poe's mistaken understanding that "statistical" probability justifies that method. Although de Queiroz and Poe assert that maximum likelihood has the power "to explain data", they do not successfully demonstrate how causal explanation is achieved or what it is that is being explained. This is not surprising, bearing in mind that what is assumed about character evolution in the accompanying likelihood model M cannot then be explained by the results of a maximum likelihood analysis. [Absolute (logical) probability; critical evidence; corroboration; explanation; falsificationism; maximum likelihood; relative (conditional) probability; severity of test; verificationism.]

Although we seek theories with a high degree of corroboration, *as scientists we do not seek highly probable theories but explanations; that is to say, powerful and improbable theories*.

> Popper (1962:58; italics in the original).

de Queiroz and Poe (2001) conjecture that my (Kluge, 1997a) understanding of Popper's (1959) degree of corroboration, $C(h,e,b)$, is incorrect, and that there is no basis for my distinguishing falsificationist and verificationist approaches to phylogenetic inference. They are unambiguous in their opinion (pp. 306):

> We argue that Popper's corroboration is based on the general principle of likelihood and that likelihood methods of phylogenetic inference are thoroughly consistent with corroboration. We also evaluate cladistic parsimony in the same context and argue that parsimony methods are compatible with Popper's corroboration . . . only if they are interpreted as incorporating implicit probabilistic assumptions. Our conclusions contradict the views of authors (e.g., Siddall and Kluge, 1997) who have attempted to justify a preference for parsimony over likelihood on the basis of Popper's concept of corroboration yet deny that parsimony methods carry probabilistic assumptions. We also argue that the likelihood approach to phylogenetic inference, which permits evaluation of the assumptions inherent in its models, is consistent with Popper's views on the provisional nature of background knowledge.

de Queiroz and Poe (p. 317; my italics) go on to conclude that

there is no conflict between parsimony and likelihood, because the general statistical perspective of likelihood—and of Popperian corroboration—subsumes all of the individual methods and models that can be applied within the context of that perspective, including those of cladistic parsimony. One of the primary advantages of adopting this perspective is that it unifies all of the various phylogenetic methods/models under a single, general, theoretical framework that allows phylogeneticists to compare those methods/models directly in terms of *their ability to explain data*. In this context, all phylogenetic methods/models are legitimate philosophically, though all have limitations, and some may explain the data better than others in particular cases. But regardless of the relationship between cladistic parsimony methods and either Popper's degree of corroboration or Fisher's likelihood, likelihood forms the basis of Popper's degree of corroboration, and likelihood methods of phylogenetic inference are fully compatible with that concept.

Thus, de Queiroz and Poe interpret degree of corroboration *solely* as a likelihood argument—no importance is attributed to *critical* evidence—and so tacitly they render Popper a verificationist. That Popper is a verificationist is extraordinary, because de Queiroz and Poe are taking a position not only on the epistemology of phylogenetic inference but also on Popper's deductive philosophy of science more generally (see epigram and the section Popper "Replies" to de Queiroz and Poe). Many philosophers and scientists have struggled to understand Popper's writings, and certainly not all agree

as to his reason, argument, and evidence. Nonetheless, even Popper's most fervent detractors agree that he sought a philosophy of science opposed to verificationism (Miller, 1999).

de Queiroz and Poe's failure to correctly interpret Popper and his falsificationist philosophy of science stems from not distinguishing their use of relative probability from his use of absolute probability. Indeed, it was that very distinction that led Popper to see that verificationism and falsificationism could not be the same.

FALSIFICATIONISM AND VERIFICATIONISM

Falsificationism is the philosophy that knowledge increases through a process of exposing false hypotheses by falsification—where the concern is that tentatively accepted hypotheses are not false. There are two kinds of falsificationism: dogmatic falsification and informed falsification (Siddall and Kluge, 1997). In turn, there are two kinds of informed falsification: methodological falsification and sophisticated falsification (sensu Lakatos, 1993). According to Kluge (1997b, 1999), and as will be reviewed below, phylogenetic systematics practiced in terms of Popperian testability is consistent with methodological falsificationism and is potentially consistent with sophisticated falsificationism, testing competing hypotheses and predictions (or retrodictions), respectively.

There are two kinds of verificationism, classical verificationism and neo-verificationism (Siddall and Kluge, 1997). The former is concerned that the accepted hypothesis be true, whereas the latter is concerned with the relative truthfulness (verisimilitude) of hypotheses, as determined by their degrees of probability. Verification and induction go hand in hand (Popper, 1959:418).

PROBABILISM

Probabilism is the doctrine that the reasonableness of hypotheses is to be judged with degrees of probability, the position of "reasonableness" being flanked by extreme optimist and skeptic positions:—"certainty can be achieved", and "probabilities cannot even be assigned" (Watkins, 1984). Which position applies to phylogenetics? If it is unrealistic to assign degrees of probability to events in history, because they are necessarily unique, then what is the basis for choosing among competing cladograms? Some have asserted repeatedly that being able to assign statistical probabilities defines the enterprise of science (e.g., Felsenstein, 1982:399, and elsewhere; see also Sanderson, 1995:300). Can phylogenetics be scientific if it is not founded on degrees of probability? An affirmative answer lies in Popperian testability (see below).

Probability has been interpreted in many different ways (e.g., Watkins, 1984). The concept of probability was born out of the desire to predict events, particularly in games of chance where wagering was involved. Classical and inverse kinds of probability represent early attempts to interpret the probability of winning, which eventually gave way to the familiar statistics of estimation and significance testing. The basis for these kinds of probability lies with the concept of relative (conditional) probability, which can be formally expressed as $p(a,b) = r$, the probability of a under condition b, where r is some fraction between 0 and 1, inclusive. Given a frequency interpretation, $p(a,b) = r$ becomes the relative frequency of a within the reference class b being equal to r, where the condition b is a random population sample.

Probability can also be stated in non-relative (non-frequentist) terms, as an absolute (logical) probability, $p(a) = r$: the probability of a, where r is the absolute value, as in the names of statements, a, b, c, ... As Popper (1983:284) clarified:

> its value r is the greater the less the statement a says. Or in other words, the greater the content of a, the smaller is the value of its absolute logical probability.

Thus, for example, a is more probable than ab, provided b does not follow tautologically from a. Absolute probability cannot be interpreted as a frequency, except in the most trivial sense.

DERIVING POPPERIAN TESTABILITY: AN EXERCISE IN FALSIFICATIONISM

Early in his career, Popper (1934, 1959:217) recognized that the *logical content* of the conjunction of two statements, $ct(xy)$, will always be greater than, or at least equal to, that of either of its components, $ct(x)$ and $ct(y)$,

$$ct(x) \leq ct(xy) \geq ct(y), \qquad (1)$$

whereas the "monotony law" of the probability calculus declares the opposite,

$$p(x) \geq p(xy) \leq p(y). \qquad (2)$$

Taking relations (1) and (2) together, Popper concluded that probability decreases with increasing content, or alternatively, *improbability* increases with increasing content, where the content of a theory (hypothesis) is equivalent to any set of statements pertaining to the proposition. A statement has logical content insofar as it prohibits or excludes something.

As Popper went on to argue, if growth of knowledge means increasing content, then surely high probability cannot be the goal of science. It also follows that a high degree of corroboration is the goal, where low probability means a high probability of being falsified.

In Popperian logic, it is the more improbable hypothesis that is valued, the proposition that has the greatest potential to be falsified; thus that which has withstood the *strongest* (*most severe*) tests is tentatively accepted.

One of the ways of defining severity of test, $S(e,b)$, is to consider content as the complement of absolute probability,

$$ct(x) = 1 - p(x), \qquad (3)$$

which can be extrapolated to

$$S(e,b) = ct(e,b) = 1 - p(e,b), \qquad (4)$$

where evidence e and background knowledge b constitute the content of the statement, and normalizing $ct(e,b)$ with the factor $(1 + p(e,b))$ leads to

$$S(e,b) = (1 - p(e,b))/(1 + p(e,b)). \qquad (5)$$

Assuming likelihood $p(e,hb) = 1$, severity of test can be generalized to

$$S(e,h,b) = (p(e,hb) - p(e,b))/(p(e,hb) + p(e,b)), \qquad (6)$$

which Popper (1962:391) defined as:

> ... the severity [*strength*] of the test e interpreted as *supporting evidence* of the theory h, given the background knowledge b.

Here, $p(e,hb)$, containing a logical conjunction (h and b juxtaposed), is the familiar probability of evidence e given *both* hypothesis h *and* background knowledge b; $p(e,b)$, containing a simple *conditional relation* (e set off by a comma from condition b), is the prior probability of evidence e given background knowledge b *alone* (without hypothesis h).

Also, severity of test can be derived by treating logical content as the reciprocal of probability,

$$ct(x) = 1/p(x). \qquad (7)$$

In whichever way severity of test is derived, the intuitively significant numerator is stated as the difference between the probability of evidence e with [$p(e,hb)$] and without [$p(e,b)$] hypothesis h, which is a *statement* of how much hypothesis h increases the probability of evidence e.

With such a definition of severity of test (relation 6), Popper (1959, 1983) formalized the explanatory power (E) of a hypothesis as

$$E(h, e, b) = S(e, h, b)$$
$$= (p(e,hb) - p(e, b))/(p(e, hb)$$
$$+ p(e, b)), \qquad (8)$$

that is, the *power* of hypothesis h to *explain* evidence e, given background knowledge b—the more severe the test by evidence e of a hypothesis h, the greater that power.

With such a measure of severity of test (relation 6) in the capacity of supporting evidence, Popper (1959, 1983) proceeded similarly to define the degree of corroboration (C) of a hypothesis as

$$C(h,e,b) = (p(e, hb) - p(e, b))/(p(e, hb)$$
$$+ p(e,b)), \qquad (9)$$

that is, the *support* (*corroboration*) provided to hypothesis h by evidence e, given background knowledge b.

Normalization factors are included in the denominators of the logical expressions defining Popperian testability, which in one way or another are intended to remove "blemishes" from the numerator. For

example, according to Popper (1983:240),

$$C(h,e,b) = (p(e,hb) - p(e,b))/(p(e,hb)$$
$$- p(eh,b) + p(e,b)), \qquad (10)$$

where (p. 242):

> ... [the denominator] makes, for every h (provided it is consistent with b) minimal and maximal degrees of corroboration equal to -1 and to the content or degree of testability of h (whose maximum is $+1$).

Severity of test (relation 6) and degree of corroboration (relation 10) are identical, except for the presence of $p(eh,b)$ in the denominator of the latter logical expression. That difference can be ignored when $p(eh,b)$ is close to 0, as it is for hypotheses with high empirical content (Popper, 1959:401).

The numerator determines the sign of $S(e,h,b)$, $E(h,e,b)$, and $C(h,e,b)$, because the denominator cannot be negative. A clear indication of the meaning of testability is the particular sign that Popper (1983:241–242) derived for certain kinds of results. Paraphrasing:

> If evidence e neither supports nor undermines hypothesis h then $S(e,h,b) = E(h,e,b) = C(h,e,b) = 0$; $S(e,h,b)$, $E(h,e,b)$, and $C(h,e,b)$ are negative when evidence e undermines hypothesis h; $S(e,h,b)$, $E(h,e,b)$, and $C(h,e,b)$ are positive when evidence e supports hypothesis h. $S(e,h,b) = E(h,e,b) = C(h,e,b) = -1$ only when evidence e absolutely contradicts hypothesis h, in light of background knowledge b. Only if $p(e,hb) = 1$, $p(e,b) = 0$, and $p(h,b) = 0$ can $S(e,h,b)$, $E(h,e,b)$, and $C(h,e,b) = +1$. That is, in order for a hypothesis to be maximally corroborable, the probability of observing the evidence e or the hypothesis h, in light of the background knowledge b, must be zero.

Although Popper did not explore at great length in one place the specifics of background knowledge b, a sample of his comments provides a basis for summarizing his view of that concept (1962:238; 1983:188):

> [what is] unproblematic ... Few parts of the background knowledge will appear to us in all contexts absolutely unproblematic, and any particular part of it *may* be challenged at any time, especially if we suspect that its uncritical acceptance may be responsible for some of our difficulties.

(1962:238):

> the old evidence, and old and new initial conditions, including if we wish accepted theories

(1962:390; 1983:236, 252):

all of those things which we accept (tentatively) as unproblematic while we are testing the theory. (b may also contain statements of the character of initial conditions.)

and (1983:244):

> knowledge which, by common agreement, is not questioned while testing the theory under investigation.

Summarizing, I consider Popper's concept of background knowledge b to comprise only currently accepted (well-corroborated) theories and experimental results that can be taken to be true while helping to guide the interpretation of evidence e on hypothesis h. Background knowledge does not include falsified theories, nor otherwise admitted false assumptions. Thus, once a theory is falsified, it can no longer serve as background knowledge. Background knowledge also does not include tautologies or null models (such as a null model of random character distribution).

de Queiroz and Poe claim there is nothing to distinguish background knowledge b from the model M assumptions of character evolution used in maximum likelihood, $p(e|M,h)$, the probability of obtaining evidence e *and* model M, given hypothesis h. However, as Siddall and Kluge (1997) pointed out, under those conditions, the model M is deterministic to the inference, in so far as it has an effect on the calculus of the method. Therefore, the model is problematic to the likelihood method, because it is deterministic to the truth of the result. Background knowledge b, such as assuming "descent, with modification" (Kluge, 1997a; see below), does not have that effect. Should "descent, with modification," prove false, minimizing unweighted steps with the parsimony algorithm would still lead to the shortest length cladogram, and the character generalities to be explained as something other than homologues (Kluge, 1997b).

Moreover, de Queiroz and Poe's claim does not address the lack of realism in model assumptions employed in maximum likelihood. For example, to assume a common mechanism (Steel and Penny, 2000), such as a homogeneity assumption of evolution in a likelihood model, knowing that the assumption is likely to be false, is not what Popper had in mind for background knowledge (Siddall and Whiting, 1999). Further, as Felsenstein (1978:408; see also 1978:409;

1981a:195; 1981b:369, 371; 1988a:123–124; 1988b:529) readily admitted,

> . . . it will hardly ever be the case that we sample characters independently, with all of the characters following the same probability model of evolutionary change

(see also Siddall and Kluge, 1999). Nor is there any realism in model assumptions that consider necessarily unique historical events as degrees of probability (Popper, 1957:106–107; Siddall and Kluge, 1997; Kluge, 1999). I remain firm in my position that there are two classes of auxiliary assumptions: background knowledge and models. When de Queiroz and Poe claim that parsimony assumes a probability model, they conflate, both philosophically and methodologically, the plausibility parsimony of Sober (1993:174; see also Penny et al., 1996) with the phylogenetic parsimony of Farris (1983,1989).

As noted above, the intuitively significant numerator, $p(e,hb) - p(eb)$, in degree of corroboration $C(h,e,b)$ declares how much hypothesis h increases the probability of evidence e, with the effect of particular evidence e varying as a function of background knowledge b. When the probability of e given both h and b is just the probability of e given b, and, $p(e,hb) = p(e,b)$, and $C(h,e,b) = 0$. Thus, for degree of corroboration to have a high likelihood, $p(e,hb)$ must be larger than $p(e,b)$, as when $p(e,b) \ll 1/2$ (Popper, 1959). As Popper (1983:238; italics in the original) emphasized:

> . . . the smaller $p(e,b)$, the stronger will be the support which e renders to h—provided our first demand is satisfied, that is, provided e follows from h and b, or from h in the presence of b . . . ,

because adding to b reduces the difference between $p(e,hb)$ and $p(e,b)$. Thus, Popper declared his minimalist philosophy when it came to auxiliary propositions, which includes model M assumptions (contra de Queiroz and Poe, 2001), just as he did when it came to ad hoc hypotheses (Popper, 1983:232).

Popperian testability also declares that likelihood $p(e,hb)$ cannot be a general measure of degree of corroboration, because degree of corroboration, $C(h,e,b)$, is strong only with critical evidence e (severe tests). Although part of the significant numerator, $p(e,hb) - p(eb)$, is the likelihood term

$p(e,hb)$, its relationship to $p(e,b)$ cannot be ignored, because $p(e,b)$ is the basis for distinguishing critical from non-critical evidence e. As Popper (1983:242; italics in the original) succinctly put it:

> . . . what about empirical evidence e which *falsifies h* in the presence of b? Such an e will make $p(e,hb) =$ zero.

Thus, de Queiroz and Poe's conjecture that $p(e,hb) = C(h,e,b)$ is false, because there is no reference to critical evidence.

POPPERIAN TESTABILITY, PHYLOGENETIC SYSTEMATICS, AND PARSIMONY

Explicating phylogenetic systematics in terms of Popperian testability, the cladogram may be considered the set of hypotheses h of interest, evidence e one or more synapomorphies, ordinarily summarized in the form of a matrix of discrete character states, accompanied by a premise of the Darwinian principles of "descent, with modification," as background knowledge b (Kluge, 1997a, 1999). Phylogenetic hypotheses have high empirical content $1- p(h,b)$, with $p(eh,b)$ being close to zero; therefore, the limit of $p(h,b)$ sets the number of possible cladograms, which is a closed hypothesis set determined by the number of terminal taxa included in an analysis. Herein lies a logical basis for increasing the number of taxa in phylogenetic systematic studies, because of the effect of that number on increasing degree of corroboration, where $C(h,e,b) = 1 - p(h,b)$ (contra Kim, 1996). As such, empirical content, $1 - p(h,b)$, provides a criterion for the evaluation of hypotheses *before* they are actually tested (Kluge, 1999).

As a rule of methodological falsification in an evaluation of scientific hypotheses, the cladogram(s) h that requires the ad hoc dismissal of the fewest falsifiers is preferred (Kluge, 1997a:88). Consider the following example, in which the hypotheses of relationships of three terminal taxa are tested with a given matrix of evidence, a particular set of synapomorphies. Given only "descent, with modification," as background knowledge b, synapomorphies characteristic of (A,B), (A,C), and (B,C) should be equally likely, all other things being equal (this logically closed hypothesis set should not be confused with the null, random distribution, of statistical inference). Thus, if a large majority of one

class of those possible synapomorphies were to be sampled by the phylogeneticist, say, the class that characterizes (A,B), then this finding may be considered improbable given the background knowledge alone, $p(e,b)$, but not under the background knowledge plus the postulated (A,B)C cladogram, $p(e,hb)$ (Kluge, 1999:429–430). The (A,B)C hypothesis is said to be corroborated to the degree that those (A,B) synapomorphies are sampled in an unbiased manner. To be sure, as explained above, for any *particular* matrix of evidence e, maximizing the likelihood $p(e,hb)$ also maximizes corroboration $C(h,e,b)$, and likelihood can be used to select among cladograms (de Queiroz and Poe, 2001).

Although phylogenetic parsimony does not directly evaluate the likelihood probability $p(e,hb)$ in Popperian testability (e.g., see Kluge, 1997a; Farris et al., 2001), that evaluation is nonetheless accomplished in two entirely different ways. As Farris (1989:107; contra de Queiroz and Poe, 2001) argued:

> A postulate of homology explains similarities among taxa as inheritance, while one of homoplasy requires that similarities be dismissed as coincidental, so that most parsimonious arrangements have greatest explanatory power.

Moreover, such a conclusion is consistent with Popper (1962:288), who pointed out that his

> ... definition $[C(h,e,b)]$ does not automatically exclude *ad hoc* hypotheses, but it can be shown to give most reasonable results if combined with a rule [such as parsimony] excluding *ad hoc* hypotheses.

In addition, there is Tuffley and Steel's (1997:599) Theorem 5: Maximum parsimony and maximum likelihood with no common mechanism are equivalent in the sense that both choose the same tree or trees. Merely assuming "descent, with modification" (Kluge, 1997a), for a given matrix of evidence e, without regard to its critical nature, explanatory power $E(h,e,b)$ is maximized for the same cladogram(s) having greatest likelihood $p(e,hb)$.

Further, as was noted above, $p(e,b)$ in the numerator of degree of corroboration $C(h,e,b)$, $(p(eh,b) - p(e,b))$, distinguishes critical from *non*-critical evidence e. Different bodies of evidence $e_1, e_2, e_3, \ldots e_n$ are regularly brought to bear on the set of possible cladograms h, or the parts of two or more different cladograms. Traditional character reanalysis is the most common basis for increasing the severity of test. This is what Hennig (1966) referred to as reciprocal clarification, and which elsewhere has been conceptualized as a never-ending cycle of research (Kluge, 1997b). In summary, most-parsimonious cladograms maximize explanatory power (Farris, 1983, 1989) and are least refuted (Kluge, 1997a), ampliative (Kluge, 1997b), and of greatest likelihood (Tuffley and Steel, 1997).

LIKELIHOOD RATIO TEST

de Queiroz and Poe (p. 319) argue the *statistical* credibility of maximum likelihood in terms of the likelihood ratio test (see Huelsenbeck and Crandall, 1997; Huelsenbeck and Rannala, 1997; Pagel, 1999). However, according to Felsenstein (1983:317; see also Huelsenbeck and Crandall, 1997), the ratios of maximum likelihoods in phylogenetic systematics are only used

> to test whether a less general hypothesis can be rejected as compared to a more general one that includes it.

The likelihood ratio test, then, does not actually test for goodness-of-fit. Rather, it is a test for the significance of how much better the fit is among alternative models. However, although one model can provide a better fit than does another, the better-fitting hypothesis does not provide a significantly good fit.

de Queiroz and Poe (p. 319) claim the likelihood ratio test has the potential to falsify model assumptions. However, the test can never be anything but an indefensible optimality criterion, because the assumptions of the model are contingencies that require testing outside the model itself (Thompson, 1975:11; see also Edwards, 1972; Goldman, 1990). The likelihood ratio test is not empirical and must not be confused with the nature of critical evidence, $p(e,hb) - p(e,b)$, in Popperian testability (contra de Queiroz and Poe). The likelihood ratio test is missing that essential ingredient of a valid scientific test, namely, empirical independence. Thus, it is ironic that the likelihood ratio test continues to be cited as the ampliative basis for maximum likelihood. In phylogenetics,

maximum likelihood would appear to have nothing to say about causal hypotheses that is not confounded by assuming what is at issue in the argument. The appearance of pursuing explanation with the likelihood ratio test is more apparent than real.

POPPER "REPLIES" TO DE QUEIROZ AND POE

de Queiroz and Poe frequently cite Popper as confirming their opinions, with several of their quotations coming from Popper's "new" Appendix *IX (Corroboration, the weight of evidence, and statistical tests) in *The Logic of Scientific Discovery* (1959: 387–419). In concluding, it seems only fitting to let Popper speak more fully on the merits of de Queiroz and Poe's claim that $p(e,hb) = C(h,e,b)$. What is to follow is the *entire* final section (*14, pp. 418–419) of Appendix *IX, in which Popper argued passionately that he is neither a verificationist nor an inductionist, and that $p(e,hb) \neq C(h,e,b)$.

It might well be asked at the end of all this whether I have not, inadvertently, changed my creed. For it may seem that there is nothing to prevent us from calling $C(h,e)$ 'the inductive probability of h, given e' or—if this is felt to be misleading, in view of the fact that C does not obey the laws of probability calculus—'the degree of the rationality of our belief in h, given e'. A benevolent inductivist critic might even congratulate me on having solved, with my C function, the age-old problem of induction *in a positive sense*—on having finally established, with my C function, the validity of inductive reasoning.

My reply would be as follows. I do not object to calling $C(h,e)$ by any name whatsoever, suitable or unsuitable: I am quite indifferent to terminology, so long as it does not mislead us. Nor do I object—so long as it does not mislead us—to an extension (inadvertent or otherwise) of the meaning of 'induction'. But I must insist that $C(h,e)$ can be interpreted as degree of corroboration only if e *is a report on the severest tests we have been able to design*. It is this point that marks the difference between the attitude of the inductivist, or verificationist, and my own attitude. The inductivist or verificationist wants *affirmation* for his hypothesis. He hopes to make it *firmer* by his evidence e and he looks out for 'firmness'—for 'confirmation'. At best, he may realize that we must not be biased in our selection of e: that we must not ignore unfavourable cases; and that e must comprise reports on our *total* observational knowledge, whether favourable or unfavourable. (Note that the inductivist's requirement that e must comprise our *total* observational knowledge cannot be represented in any formalism. It is a non-formal requirement, a condition of adequacy which must be satisfied if we wish to interpret $p(h,e)$ as degree of our imperfect knowledge of h.)

In opposition to this inductivist attitude, I assert the $C(h,e)$ must not be interpreted as the degree of corroboration of h by e, unless e reports the results of *our sincere efforts to overthrow h*. The requirement of sincerity cannot be formalized—no more than the inductivist requirement that e must represent our total observational knowledge. Yet if e is *not* a report about the results of our sincere attempts to overthrow h, then we shall simply deceive ourselves if we think we can interpret $C(h,e)$ as degree of corroboration, or anything like it.

My benevolent critic might reply that he can still see no reason why my C function should not be regarded as a positive solution to the classical problem of induction. For my reply, he might say, should be perfectly acceptable to the classical inductivist, seeing that it merely consists in an exposition of the so-called 'method of eliminative induction'—an inductive method which was well known to Bacon, Whewell, and Mill, and which is not yet forgotten even by some of the probability theorists of induction (though my critic may well admit that the latter were unable to incorporate it effectively into their theories).

My reaction to this reply would be regret at my continued failure to explain my main point with sufficient clarity. For the sole purpose of the elimination advocated by all these inductivists was to *establish as firmly as possible the surviving theory* which, they thought, must be the *true* one (or perhaps only a *highly probable* one, in so far as we may not have fully succeeded in eliminating every theory except the true one).

As against this, I do not think that we can ever seriously reduce, by elimination, the number of the competing theories, since this number remains always infinite. What we do—or should do—is to *hold on, for the time being, to the most improbable of the surviving theories* or, more precisely, to the one that can be most severely tested. We tentatively 'accept' this theory—but only in the sense that we select it as worthy to be subjected to further criticism, and to the severest tests we can design.

On the positive side, we may be entitled to add that the surviving theory is the best theory—and the best tested theory—of which we know.

SUMMARY

My particular disagreements with de Queiroz and Poe concerning maximum likelihood (L) and Popperian degree of corroboration (C) can be summarized succinctly:

$$L(h, e) = p(e \mid M, h), \tag{11}$$

whereas

$$C(h,e,b) = p(e, hb) - p(e, b)/\cdots, \tag{12}$$

and therefore

$$L(h,e) \neq C(h,e,b). \tag{13}$$

This conclusion results from the fact that M includes counterfactual assumptions, which are unlikely to be true and, which are excluded from b, and because $L(h,e)$ maximizes the likelihood function with *non*-critical evidence, just $p(e \mid M, h)$, whereas $C(h, e, b)$ maximizes the likelihood function with *critical* evidence, $p(e, hb) - p(e, b)$.

Like Popper, my general disagreement with likelihood inference is that the research program is, by definition, verificationist; that is, truth is sought inductively with degrees of probability. Falsificationism seeks explanation deductively with potential falsifiers, which obtains in phylogenetic systematics when the most-parsimonious, least disconfirmed, cladogram is sought, because it maximizes the explanation of synapomorphies in terms of inheritance.

REFERENCES

DE QUEIROZ, K., AND S. POE. 2001. Philosophy and phylogenetic inference: A comparison of likelihood and parsimony methods in the context of Karl Popper's writings on corroboration. Syst. Biol. 50:305–321.

EDWARDS, A. W. F. 1972. Likelihood. Cambridge Univ. Press, Cambridge [1992 reprint].

FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. Pages 7–36. *In* Advances in cladistics (N. I. Platnick and V. A. Funk, eds.), volume 2. Columbia Univ. Press, New York.

FARRIS, J. S. 1989. Entropy and fruit flies. Cladistics 5:103–108.

FARRIS, J. S., A. G. KLUGE, AND J. M. CARPENTER. 2001. Popper and likelihood versus "Popper*". Syst. Biol. 50:438–444.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

FELSENSTEIN, J. 1981a. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. 16:183–196.

FELSENSTEIN, J. 1981b. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

FELSENSTEIN, J. 1982. Numerical methods for inferring evolutionary trees. Q. Rev. Biol. 57:379–404.

FELSENSTEIN, J. 1983. Methods for inferring phylogenies: A statistical view. Pages 315–334. *In* J. Felsenstein (ed.), Numerical taxonomy. Springer-Verlag, New York.

FELSENSTEIN, J. 1988a. The detection of phylogeny. Pages 112–127. *In* Prospects in systematics (D. L. Hawksworth, ed.). Systematics Assoc., Clarendon Press, Oxford.

FELSENSTEIN, J. 1988b. Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. 22:521–565.

GOLDMAN, N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson Process Model of DNA substitution and to parsimony analyses. Syst. Zool. 39:345–361.

HENNIG, W. 1966. Phylogenetic systematics. Univ. of Illinois Press, Chicago.

HUELSENBECK, J. P., AND K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. 28:437–466.

HUELSENBECK, J. P., AND B. RANNALA. 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. Science 276:227–232.

KIM, J. 1996. General inconsistency conditions for maximum parsimony: Effects of branch lengths and increasing numbers of taxa. Syst. Biol. 45:363–374.

KLUGE, A. G. 1997a. Testability and the refutation and corroboration of cladistic hypotheses. Cladistics 13:81–96.

KLUGE, A. G. 1997b. Sophisticated falsification and research cycles: Consequences for differential character weighting in phylogenetic systematics. Zool. Scr. 26:349–360.

KLUGE, A. G. 1999. The science of phylogenetic systematics: Explanation, prediction, and test. Cladistics 15:429–436.

LAKATOS, I. 1993. Falsification and the methodology of scientific research programmes. Pages 91–196. *In* Criticism and the growth of knowledge (I. Lakatos and A. Musgrave, eds.). Cambridge Univ. Press, London.

MILLER, D. 1999. Being an absolute skeptic. Science 284:1625–1626.

PAGEL, M. D. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

PENNY, D., M. D. HENDY, P. J. LOCKHART, AND M. A. STEEL. 1996. Corrected parsimony, minimum evolution, and Hadamard conjunctions. Syst. Biol. 45:596–606.

POPPER, K. 1934. Logik der Forschung. Springer Verlag, Vienna.

POPPER, K. 1957. The poverty of historicism. Routledge and Kegan Paul, London.

POPPER, K. 1959. The logic of scientific discovery. Harper and Row, New York.

POPPER, K. 1962. Conjectures and refutations: The growth of scientific knowledge. Routledge and Kegan Paul, London.

POPPER, K. 1983. Realism and the aim of science. Routledge, London. [1992 reprint]

SANDERSON, M. J. 1995. Objections to bootstrapping phylogenies: A critique. Syst. Biol. 44:299–320.

SIDDALL, M. E., AND A. G. KLUGE. 1997. Probabilism and phylogenetic inference. Cladistics 13:313–336.

SIDDALL, M. E., AND A. G. KLUGE. 1999. Letter to the editor [evidence and character independence]. Cladistics 15:439–440.

SIDDALL, M. E., and M. F. WHITING. 1999. Long branch abstractions. Cladistics 15:9–24.

SOBER, E. 1993. Philosophy of biology. Westview Press, San Francisco, California.

STEEL, M., and D. PENNY. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol. Biol. Evol. 17:839–850.

THOMPSON, E. A. 1975. Human evolutionary trees. Cambridge Univ. Press, Cambridge.

TUFFLEY, C., AND M. STEEL. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull. Math. Biol. 59:581–607.

WATKINS, J. 1984. Science and scepticism. Princeton Univ. Press, Princeton, New Jersey.