On Probability and Systematics

1. Introduction

Within the field of phylogenetic systematics, an ongoing debate has revolved around the appropriate choice of methodology for cladistic analysis, i.e., the construction of phylogenetic trees and inference of ancestral states. A recent paper by Mark Siddall and Arnold Kluge (Siddall and Kluge, 1997) highlights this controversy in theoretical systematics. Siddall & Kluge's paper has generated a lot of reaction, and characterizes the kinds of arguments which have strongly influenced biologists' choice of methodology. Siddall & Kluge advocate a privileged status for a phylogenetic technique called parsimony, to the exclusion of other, statistically based, phylogenetic methods. The present paper will address some of the debates of their argument. Rather than try to address every point made by Siddall & Kluge, I draw out two underlying general lines of argument that are representative of a particular school of thought in phylogenetics. (This school of thought is typically referred to as the Cladistic school.) The two lines of argument that I identify are what I term (i) the *argument from falsificationism*, and (ii) the *argument from probability*. The first of these has been addressed elsewhere both by philosophers and biologists, and will merely be commented upon below. The argument from probability, though, is the primary focus of this paper. I show that the argument from probability, as invoked by Siddall & Kluge, is ambiguous, *e.g.*, between metaphysical and epistemic possibility. Upon disambiguation, the argument from probability is either invalid, unsound, or simply misses the intended target. In working through this disambiguation, I precisely identify and clarify Siddall & Kluge's concerns, and show that statistical phylogenetic techniques ought not be considered problematic for the reasons cited by Siddall & Kluge.

2. Historical Context

To explain the importance of addressing Siddall & Kluge's arguments, I will first outline the context in which they appear. This requires a brief review of the history of modern systematics.

Today, almost all systematists are *phylogenetic systematists*, or *cladists*. Broadly speaking, this simply means that systematists share a common goal of building *phylogenetic trees*. A phylogenetic tree is a way of displaying the evolutionary relationships of the groups being studied. The kinds of groups whose evolutionary relations can be mapped onto a phylogenetic tree range from HIV variants to species to orders to kingdoms of organisms, and are generically referred to as *taxa*.

Though it is now (for the most part) uncontroversial that the broad goal of building trees is to display the evolutionary relations of the taxa in question, it was not always so. David Hull has catalogued the emergence of modern systematic biology (Hull, 1988). The cladistic school arose in the late 1960's/early 1970's and challenged what was known as *numerical phenetics*. *Pheneticists* advocated using similarity algorithms as a method of classification of taxa. This methodology involved coding various characters of the taxa in question (e.g., morphological traits such as fur length), and then using a similarity matrix to cluster those groups that were most similar (see *fig. 1b*). Notably, numerical pheneticists did *not* think that evolutionary relationships of these groups was something that could be validly inferred, and strove for what they claimed would be an objective classification based on observable character states (Sokal and Sneath, 1963).

Phylogeneticists argued that no such objective classification was possible based on the inductive methods of phenetic numerical taxonomy, and that inferences of evolutionary relationships of taxa could be justified. Phylogeneticists followed Willi Hennig in taking the aim of systematics to be the construction of phylogenetic trees that reflected nested sets of sister-taxa, that is, taxa that are descended from a common ancestor (see *fig. 1a*) (Hennig, 1979). These sets of sister-taxa are clades, hence the term *cladistic analysis*. Phylogenetic analyses produce *cladograms* which display evolutionary relationships, but do not necessarily convey any objective similarity between the taxa. This reflects the phylogenetic claim that similarity alone does not necessarily correlate to evolutionary relatedness. Not surprisingly, phenetic and phylogenetic analyses of data produce much different results (see *fig. 1*).

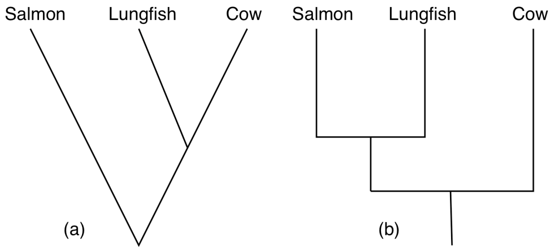

Figure 1. A phylogenetic classification (a) groups taxa by evolutionary relatedness, as defined by most recent common ancestor. The result is a *cladogram* displaying nested sets of clades. Phenetic classifications (b) produce phylograms which group taxa according to similarity indexes, and do not display evolutionary relationships

The phylogenetic technique of choice among cladists was broadly known as parsimony (Hull, 1988; Sober, 1988). Though there are some variants of parsimony, for the purpose at hand these can be ignored. As the name would suggest, parsimony works by picking out the phylogenetic tree that requires the least number of evolutionary events yet is consistent with the observed data.[1] Many leading cladistic theorists initially justified parsimony techniques by explicitly appealing to Popperian falsificationism (Eldredge and Cracraft, 1980; Farris, 1983; Kluge, 1997a; Kluge, 1997b). The most parsimonious phylogenetic tree was said to be the most corroborated phylogenetic tree, and was understood as a bold hypothesis of the evolutionary relationships between the taxa being studied. This phylogenetic tree was subject to being falsified if the discovery or addition of new characters into the study revealed a more parsimonious phylogenetic tree.

---

[1] The data that is observed is the distribution of characters across the groups being studied. Characters can be anything from nucleotide sequences to morphological characteristics to behavioral traits.

The reconstruction of phylogeny using parsimony involves the construction of a tree-like (*i.e.,* bifurcating) model to represent a section of the actual branching patterns of a historical lineage (*e.g.*, a new branch may represent a speciation event). Note the importance of the distinction between the 'true tree' and a 'phylogenetic tree.' The 'true tree' is the actual historical lineage of life, (or part of that actual historical lineage). The 'phylogenetic tree', on the other hand, is a model that systematists consider to be a hypothesis about the structure of the true tree. Parsimony provides a way to both construct and evaluate different models/hypotheses of a section of the actual historical lineage. This distinction will turn out to be very important, and will be revisited below.

In the late 1970's, phylogeneticist Joseph Felsenstein discovered that trees of a certain shape were likely to cause a systematic error in the parsimony method (Felsenstein, 1978). This problem was called long-branch attraction (after the shape of the tree described by Felsenstein; see *fig. 2*). In phylogenetic trees, branches connect nodes to other nodes or taxa. The length of these branches can be used to represent the amount of evolutionary change along a branch. In long-branch attraction, the taxa at the ends of long branches are mistakenly grouped together by parsimony instead of with the groups with which they actually share a more common recent ancestor. Cases in which this error can occur have been dubbed the Felsenstein Zone

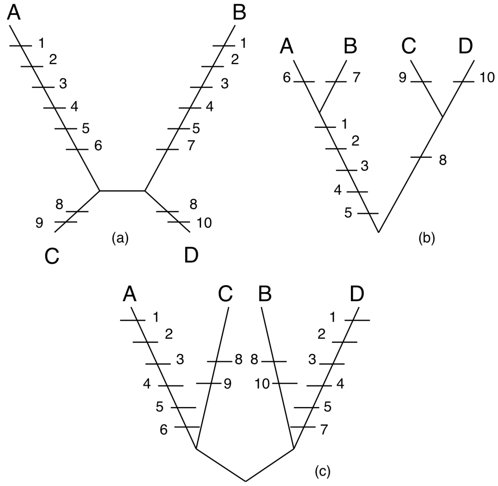

Figure 2. (*a*) represents the actual pattern of evolution, (*i.e.,* the actual historical lineage of A, B, C and D), with the branch lengths indicating amount of evolutionary change. This tree is unrooted, *i.e.*, it is not time directional. The phylogenetic trees at (b) and (*c*) are rooted, with extant taxa at the labeled tips. (*b*) is a parsimony analysis of the data from (*a*), whereas (*c*) is a ML analysis of the same data. Note that the evolutionary relationships displayed in the parsimony phylogenetic tree are not isomorphic with the actual evolutionary relationships.

(Huelsenbeck and Hillis, 1993; Huelsenbeck, 1997). Long-branch attraction is not merely a problem at an operational level, but also presents a challenge to the conceptual underpinnings of parsimony. Recall that some cladists couched the justification for parsimony in falsificationist terms. That is, as more data (in the form of characters) are added or discovered, the most parsimonious phylogenetic trees are subject to being falsified in favor of more parsimonious phylogenetic trees. These new phylogenetic trees, then, are held up as bold hypotheses and either corroborated or rejected in favor of ever more parsimonious trees which are subject to being tested. Felsenstein, though, showed that as more characters are added, parsimony techniques become *more subject* to making a long-branch attraction error. This is because as more characters are added, it raises the possibility that there will be a long branch in the tree.

Felsenstein proposed using a new method to build phylogenetic trees: maximum-likelihood methods (ML) (Felsenstein, 1981). ML picks out the phylogenetic tree that has the highest likelihood value (conditional on the data). It has been shown that ML methods are not subject to long-branch attraction (see *fig. 2c*) (Huelsenbeck and Hillis, 1993; Hillis et al., 1994; Swofford et al., 1996; Huelsenbeck and Rannala, 1997). Despite Felsenstein's hopes, ML methods were not welcomed by the leading cladistic theorists (Felsenstein, 2001). A split formed in the systematics community between those that advocated using parsimony techniques exclusively, and those that advocated using statistical techniques such as ML. The former group appropriated the name *Cladists* (whom I call capital 'C' cladists); the latter group identified themselves as *statistical phylogeneticists*. Very recently, a new group has emerged within statistical phylogenetics: *Bayesians* (Lewis and Swofford, 2001; Huelsenbeck et al., 2002). As the name would imply, Bayesians use Bayesian phylogenetic techniques to construct phylogenetic trees. Though some non-Bayesian statistical phylogeneticists are skeptical of the Bayesian techniques, they do consider Bayesians to be fellow statistical phylogeneticists. Statistical phylogeneticists typically consider parsimony methods to be just another statistical method – one that can be used when long-branch attraction is not an issue. Cladists, on the other hand, do not consider statistical techniques to be valid forms of phylogenetic inference, and some also question whether statistical methods conform to the 'scientific method' (i.e., a particular form of Popperian falsificationism).

The theoretical debates that occur within systematics can often be daunting and inaccessible to the non-specialist, though the consequences of the debate are extremely relevant and important to their work. The products of systematics are widely used in the field of biology, and biologists must choose which method to use (or use extra resources by utilizing multiple methods). A recent Cladist article by Siddall & Kluge (Siddall and Kluge, 1997) is representative of the kinds of arguments underlying Cladist assertions against statistical phylogenetics. The Siddall & Kluge paper is very complex and dense. There is neither room nor do I think it would be particularly productive to go through their paper point by point. Instead I precisely identify two of the underlying arguments as I understand them. In so doing, I have identified specific points of conflict and confusion between the Cladists and the statistical phylogeneticists. By carefully analyzing one of these arguments, I show that upon disambiguation the argument does not tell against using statistical techniques in phylogenetics.

3. The Argument From Falsificationism

Broadly speaking, Siddall & Kluge have two main lines of argument implicit in their paper: (i) *the argument from falsificationism*; and (ii) *the argument from probability*. I will explore the argument from probability in more detail in the following section. First, though, some brief comments on the argument from falsificationism.

The argument from falsificationism can be schematized as follows:

F1 The desired scientific methodology is (some kind of) Popperian falsificationism.
F2 Parsimony is a phylogenetic method consistent with Popperian falsificationism.
F3 Statistical phylogenetic techniques are not consistent with Popperian falsificationism.
F4 The only available phylogenetic technique which embodies a falsificationist methodology is parsimony.
F5 Therefore, parsimony is the only phylogenetic technique that conforms to the desired scientific methodology.

Siddall & Kluge's argument from falsificationism has spawned an ongoing debate in the systematics literature, *e.g.,* the recent exchange in *Systematic Biology* between Kluge and DeQueiroz & Poe (de Queiroz and Poe, 2001; Kluge, 2001). Statistical phylogeneticists have tended to argue against Siddall & Kluge in one of two ways. The first is to argue that statistical phylogenetic techniques do, in fact, conform with Siddall & Kluge's characterization of a falsificationist scientific methodology (*i.e.*, to deny premise F3). The other strategy has been to argue that Siddall & Kluge are offering a mistaken interpretation of Popperian falsificationism which does not qualify as a criterion by which statistical methods ought to be judged (*i.e.*, to argue that Siddall & Kluge mischaracterize falsificationism in F1).

There is a further question of how the falsificationism espoused by Siddall & Kluge resembles that which has been discussed in the philosophical literature (see (Hull, 1983) for an earlier treatment of Cladist characterizations of falsificationism). There do seem to be at least some important differences, *e.g.,* there appears to be some incongruence over the treatment and classification of Fisherian statistics (Gillies, 1990; Urbach, 1991; Siddall and Kluge, 1997), though this is a discussion that must be left for another day.

Philosophers have also evaluated Popperian falsificationism (though only a few have done so in the context of systematics (Hull, 1983; Sober, 1983; Hull, 1988; Sober, 2000)). Most contemporary philosophers of science are critical of the idea that falsificationism is the only acceptable scientific methodology (Kuhn, 1970; Lakatos, 1970; Grunbaum, 1976; Kitcher, 1982; Kuhn, 1996; Salmon, 1998; Sober, 2000). Indeed, some philosophers have gone so far as to question whether falsificationism is even a very good scientific methodology (Howson and Urbach, 1993). The reasons for these objections are many and varied, and I will not rehash them here. Suffice to say that most philosophers of science would be mildly surprised that very few attempts have been made to deny premise F1 of the argument from falsificationism. Though I think this strategy might prove fruitful, it is not within the aims of the present paper to pursue this strategy.[2]

4. The Argument from Probability

Siddall & Kluge also argue for the privileged status of parsimony techniques on the basis of their interpretation of probability. What I term their *argument from probability* runs roughly as follows:

P1 Phylogenetic trees are unique historical entities.
P2 Probabilities cannot be assigned to unique historical entities.
P3 Therefore, phylogenetic trees are not the kinds of things to which probabilities can be assigned.
P4 Statistical methods claim to assign probabilities to phylogenetic trees.
P5 Therefore, statistical methods are not applicable to the building of phylogenetic trees.

In what follows, I examine each of the premises identified above, and show that each is either ambiguous or false. Upon disambiguation, it is evident that the argument from probability is either invalid, unsound or simply misses its intended target.

4.1. Concerning P1: Phylogenetic Trees Are Unique Historical Entities

Recall the distinction between the *true tree* and *phylogenetic trees*. The term 'true tree' refers to either the actual historical lineage of life, or to a segment of that lineage. Which sense of

---

[2] Note that by denying Popperian falsificationism one is *not* denying that testing hypotheses and (possibly) proving them false is an important component of scientific examinations. One must not confuse falsificationism with any act of falsifying hypotheses; to do so is to get caught in the fallacy of persuasive terminology. (Thanks to Michael Jubien for letting me borrow this phrase.)

'true tree' is being used is usually clear from the context in which it is used, and this is generally not problematic. Unless otherwise noted, the term 'true tree' as I use it should be understood as referring to 'some actual segment of the historical lineage of life'.

Systematists also talk about phylogenetic trees. 'Phylogenetic tree' should be understood as a hypothesis about a particular segment of the actual historical lineage of life. So to speak of the true tree is to talk about an actual historical entity that is a part of the unique historical lineage of life, whereas to speak of a phylogenetic tree is to talk about a hypothesis about the true tree.

To fully grasp the weight of Siddall & Kluge's argument, it is essential to comprehend their notion of actual and possible, combined with their acceptance of historical lineages as individuals[3] (1997, p. 314-15):

> There remains considerable confusion in comparative biology concerning universals and particulars. A simple question-answer exchange between a probabilist and a historian illustrates how easy it is to conflate the two.
> *Probabilist*: "What is the chance of life evolving on earth?"
> *Historian*: "Chance? It simply did."
> *Probabilist*: "What is the chance that life has evolved, or could evolve, elsewhere in the universe?"
> *Historian*: "None."
> *Probabilist*: "Don't we have a good idea of the physical and chemical conditions necessary for life on earth, the number of appropriate stars and M-class planets, and, from that, would you not agree that we can predict the likelihood of there being life elsewhere?"
> *Historian*: "Certainly not. Of course the answer might have been yes, if I had understood your question to mean a *kind* of life. Obviously, your question is metaphysical, as opposed to scientific."

There are several points worth drawing out of this dialogue.[4] First is that Siddall & Kluge are taking life on earth to be a unique historical entity. Secondly, Siddall & Kluge want to

---

[3] Thinking of historical lineages, in general, as individuals is an extension of thinking of species as individuals. See (Ghiselin, 1974; Hull, 1976) for more on the thesis that species are individuals, and (Baum, 1998) for extending this thesis to any historical lineage. Though recent developments in metaphysics suggest some revisions to this thesis are necessary, these revisions are irrelevant to the topic at hand (Jubien, 2001). Broadly speaking, I accept as correct the thesis that historical lineages ought to be thought of as individuals.

[4] A side note worth mentioning is that even if we accept Siddall & Kluge's presuppositions, it still seems possible that life could have evolved, or could evolve, elsewhere in the universe. Life on earth might have started due to a 'seed' from a different part of the universe, and we may yet send life to other planets where it

dismiss 'chance' as being relevant to the generation of life on earth. Finally, Siddall & Kluge are claiming that discussions of possibility of things like historical individuals are discussions of metaphysical possibility. Taken together, Siddall & Kluge are denying the chance, or possibility, that life might not have evolved on earth.

But certainly we can imagine a possible world in which life failed to evolve on earth, which suggests that this is a metaphysical possibility. Are Siddall & Kluge denying this, and asserting instead that what is possible is only that which is actual? If so, fine; it is useful in and of itself to identify and clarify this presupposition of theirs, even if it conflicts with most standard philosophical (and, presumably, scientific) usage in which the actual is understood to be one among many possibilities. We can simply read this as a controversial claim, and, thus, as a source of controversy and confusion among Siddall & Kluge's critics. Note that we are still speaking in terms of metaphysical possibility, *i.e.,* what kinds of things are possible in the world irrespective of our knowledge of those things. Let's look at how Siddall & Kluge want to apply the notions developed in the above dialogue.

Siddall & Kluge argue that there is only one actual historical lineage, and want to dismiss all talk of possible trees (1997, p. 317):

> For frequency probability to apply to phylogeny there has to be a set of *simultaneously* possible trees, but if only one tree can be "true" then all others are necessarily false.

This too, though, is a metaphysical claim, and pertains to a question of what kinds of things are possible irregardless of our knowledge. When biologists speak of 'possible trees', however, the claim is not metaphysical, but epistemic; *i.e.*, biologists are not denying that there is a unique historical lineage, rather they are simply claiming that given the state of our knowledge, a range of models or hypotheses about the actual lineage are consistent with

---

might continue to evolve. If either of these cases turn out to be true, then life on earth is merely a part of life in the universe.

what we know of that lineage. So given our epistemic position, discussion of possible trees can be understood as an epistemic statement, *not* as the metaphysical claim that Siddall & Kluge are trying to make it into.

With regard to epistemic possibility, the concern is over things like a range of hypotheses about a single event. At times, Siddall & Kluge seem to recognize this (1997, p. 313-14):

> No one disputes what the alternative hypotheses are in phylogenetics. That is, for *N* taxa there are exactly $(2N\text{-}3)!/\mathrm{N}^{N\text{-}2}(N\text{-}2)!$ possible bifurcating cladograms, all of which are capable of explaining observed character state distributions. These trees, then, comprise part of the premise for any phylogenetic analysis irrespective of method.

This is *not* a metaphysical problem; the problem is that the state of the actual event is not known, and there are a range of possible hypotheses or models that describe the event in question and are consistent with our beliefs about that event. The question for systematists, of course, is how to evaluate which of these explanations is best supported or justified. Philosophers call this epistemic possibility, and it is not problematic from a metaphysical perspective (Kripke, 1980; Jubien, 1997). When faced with epistemic possibility, there are many alternative methods available to evaluate the competing hypotheses. Some of these methods are scientific. Falsificationism is one proposed scientific method (Popper, 1959a; Kluge, 1997b). With regard to systematics, this is manifest (or so it is claimed) in parsimony techniques. There are other scientific methods of evaluating competing hypotheses, including statistical methods. Bayesian posterior probabilities can be assigned to competing hypotheses, providing scientists with an evaluative tool while satisfying the axioms of probability. Scientists can also evaluate competing hypotheses using likelihood values, though these are not, strictly speaking, probabilities.

Despite recognition of the problem of epistemic possibility, Siddall & Kluge confuse issues by not carefully distinguishing between epistemic and metaphysical possibility in systematics (1997, p. 314):

> The problem with the verificationist program is that it denies nothing. … Verificationist approaches to phylogenetics, like maximum likelihood, suffer from this failure as well, because all trees are assigned a non-zero probability, and yet no more than one tree actually can be correct – thus the probabilities are not *explanatory*

Except, of course, that not all phylogenetic trees are assigned the *same* non-zero probability. As described here by Siddall & Kluge, the phylogenetic tree with the highest probability might *reasonably* also be considered the best explanation. But the main problem here, as elsewhere, is that Siddall & Kluge fail to distinguish between the true tree and phylogenetic trees. In the passage above, it is extremely unclear which sense of 'tree' they are using; indeed, on one reading they seem to go from one sense to the other in the same sentence. Siddall & Kluge are also confusing probabilities being assigned to events or individuals with probabilities being assigned to beliefs or hypotheses. This is an important distinction which will be discussed in more detail below.

The point of all this is to highlight Siddall & Kluge's failure to distinguish between when biologists are speaking of the true tree and when they are referring to phylogenetic trees. While the true tree is a unique historical entity, phylogenetic trees are not historical entities but models or hypotheses of the true tree.[5] Recall premise P1 of the argument from probability: *Phylogenetic trees are unique historical entities.* P1 is false, and needs to be revised. I propose the following revision, which disambiguates some of the concepts discussed above:

P1a The true tree (*i.e.*, the actual historical lineage of life) is a unique historical entity which is unobservable and, in practice, unknowable.
P1b Phylogenetic trees are hypotheses about the structure of the true tree.

---

[5] In fact, most biologists acknowledge that the true tree is likely not shaped like a bifurcating tree, and includes events such as speciation *via* hybridization and symbiosis. Bifurcating phylogenetic trees of the sort constructed by systematists are generally recognized to fail to represent these kinds of events and, thus, are *not* isomorphic to the true tree in this regard.

4.2. Concerning P2: Probabilities cannot be assigned to unique historical entities

Siddall & Kluge purport to be concerned with a frequentist interpretation of probability.[6] In particular, with whether frequentist interpretations of probability can be assigned to historical entities. Rather than historical entities, I am going to discuss singular events. This is a broader category, but includes historical entities. So the question becomes whether a frequency probability can be assigned to a singular event. Siddall & Kluge are right that it is not appropriate to assign a frequentist interpretation of probability to a singular event. However, there are other interpretations of probability available which do allow for such an assignment. To see why this is so it is instructive to briefly review these different interpretations of probability, and how each treats singular events.

Before proceeding, a useful distinction needs to be made. Similar to the distinction made above between metaphysical and epistemic possibility, so too is there a distinction between metaphysical and epistemic interpretations of probability.[7] Metaphysical interpretations of probability are those which take probability to be a thing of the world which exists independent of us, and which may also be discovered by us. Epistemic interpretations of probability, on the other hand, take probabilities to be reflections of degrees of belief about a proposition of some event or object of the world. So epistemic probabilities, then, do not exist in the world independently of our beliefs. A brief example can help draw out the importance of making this distinction.

---

[6] I admit to some confusion over their precise target, given their identification of frequentist probability with Bayes theorem (Siddall & Kluge, 1997, p. 313). Of course, Bayes theorem, in and of itself, does not indicate any conflict with a frequentist interpretation of probability. Bayes theorem is a consequence of the axioms of probability, and can be understood as a way to evaluate relative frequencies. That said, Bayes theorem is typically not associated with a frequentist interpretation, rather, it generally conveys a subjectivist interpretation of probability (Gillies, 2000; Howson and Urbach, 1993).

[7] This distinction is also sometimes described as one of objective and subjective interpretations of probability.

Suppose, for example, that I had a coin which was known to be biased, though the direction of that bias was unknown. Suppose, too, that I asked both a metaphysical and an epistemic probabilist what the probability was that the coin would land 'heads' upon flipping. The metaphysical probabilist would respond with something like, "if by 'probability' you mean a metaphysical probability, then all I can say of the biased coin is that the probability of that coin landing heads in any given coin flip is *not* 50%. The actual metaphysical probability of the coin landing heads is something that we can discover upon experiment and observation; but, given that the coin is biased, we know the probability cannot be 50%." The epistemic probabilist, on the other hand, might respond to the same question as follows, "if by 'probability' you mean epistemic probability, I have no reason for believing that the coin is biased either towards heads or tails, so the only justified degree of belief is that it is equally likely to be biased in either direction, and, thus, I can contingently assign a 50% probability to the proposition that the coin will land heads. Upon experiment and observation, we will be justified in adjusting our degree of belief accordingly." So if one is not careful to be precise about what kind of interpretation of probability is being discussed, there is great danger of mischaracterizing assignments of probability. In the example just given, both parties were right to gently chastise my ambiguous phrasing of the question, as the divergent answers given turned on which classification of probability was being assumed.

Siddall & Kluge are most concerned about the use of frequentist interpretations in phylogenetics (1997, p. 332, emphasis added):

> Take, for example, the gambler's fallacy: Roberto Alomar is batting 0.300. He comes to bat three times in a game and fails to get a hit. … Our objective probabilist [a frequentist], like the likelihoodist, … asserts that, because he is batting 0.300, he still has only a 30% chance of getting a hit, but this too fails to take into account the full scope of knowledge. In the first place, because Alomar failed to get a hit in his last three times at bat, he is actually batting 0.297; the probabilities have changed, because they are historically contingent phenomena. *More to the point, Alomar*

> *either will or he will not get a hit and there is no probability that can be assigned to that one event: betting on one event alone is foolish.*

A real life example of this can be used to draw out the principles of the different interpretations of probability. In the 2002 World Series one of the teams involved, the San Francisco Giants, had a player who is generally regarded as the best hitter in the game, Barry Bonds. Furthermore, Bonds was in the midst of what many baseball experts considered to be one of the greatest seasons in the history of the game. In response to his high level of play, opposing teams employed atypical strategies when Bonds came to bat. Bonds was intentionally walked in situations that generally do not call for such a strategy. In the World Series, the opposing team's manager (Mike Scioscia of the Anaheim Angels) employed this strategy to such an extreme that it became rather controversial.

So we now have a real life example of the kind of situation that Siddall & Kluge describe in the passage above. Faced with a decision of whether or not to pitch to Bonds in a particular at-bat, we can now ask how Scioscia might have considered his strategy on different interpretations of probability. The different interpretations that I will consider are the frequency, propensity and Bayesian interpretations of probability.

Frequency interpretation of probability

The frequency interpretation of probability defines probabilities as the long-run relative frequency of an event *m* occurring in a sequence of *n* cases where *n* is very large (or infinite)(Popper, 1959b).[8] Frequency probabilities can be discovered by observing the relative frequency of an event occurring in an observed number of cases, and then idealizing from this relative frequency to a long-run relative frequency. Frequentist interpretations of probability are metaphysical, *i.e.*, they assert that probability is a thing in the world which can

---

[8] Whether the sequence is extremely long or infinite, or actually or potentially existing, is characteristic of different versions of the frequency interpretation of probability. Though these are important distinctions, for the purpose at hand they can be ignored, and I will typically speak of these sequences as though they need only be very large.

be discovered. Most relevantly, a frequency probability simply *is* the relative frequency of an event in the long-run or infinite sequence of cases.

As Siddall & Kluge point out, any particular at-bat is a singular event which is part of the sequence of cases in which the event of getting-a-hit either occurs or does not occur. A singular event, in itself, does not make up a long-run sequence, nor can a long-run relative frequency be extrapolated from a singular event.[9] It is for this reason that frequency probabilities are not assigned to singular cases. So the frequency probability of Bonds getting-a-hit simply *is* the relative frequency of that event occurring in the appropriate sequence. But herein lies the problem. Each singular at-bat is a member of many sequences each of which may have a different relative frequency of Bonds getting-a-hit. Which of these sequences has the relative frequency of interest and relevance? Scioscia was faced with just this problem whenever Bonds came up to bat in the 2002 World Series. The relative frequency of Bonds getting-a-hit was different depending upon which conditions defined the sequence of interest (this sequence is sometimes referred to as the *reference class*). Consider the different conditions (and resultant sequences) which Scioscia might have considered:[10] Bonds' batting average could be considered over the course of his career, as a member of the Giants, in a particular baseball park, against left- or right-handed pitching, over his career playoff games, in the 2002 playoffs, etc. In fact, the relative frequency of getting-a-hit in these different reference classes or sequences diverge substantially. The problem facing Scioscia is how to determine which of the possible sequences is most relevant for a particular at-bat.

---

[9] Or, alternatively, the only long-run relative frequency which could be extrapolated is 1 or 0, which is rather uninformative.

[10] For simplicity's sake, I am using Bonds' batting average as a proxy for the relative frequency of Bonds getting-a-hit in a given sequence.

Much to the chagrin of Scioscia (and to the joy of Giants fans), Bonds' relative frequency of getting-a-hit in the 2002 World Series was much higher than in other playoff series in which he had participated. Were Scioscia a frequentist, he might have explained this in one of two ways. Scioscia might have argued that the only relevant reference class for Bonds' 2002 playoff at-bats were, in fact, Bonds' 2002 playoff at-bats. Alternatively, Scioscia might have argued that Bonds' at-bats during the 2002 playoffs made up only a very small sample size of possible cases, and that it is not at all surprising for the relative frequency of an event occurring in a small subset of a large (or, worse, infinite) sequence to diverge widely from the relative frequency of that event in the larger sequence.

In the first case the sequence is too small to qualify as a sequence from which a frequency probability can be meaningfully derived. The conditions that define the small sequence are too confining to generate a meaningful long-run relative frequency. In the second case, we are faced with a related problem. What relevancy does a frequency probability have to any given subset of the long-run sequence? If the relative frequency of an event occurring in any observable subset might diverge widely from the relative frequency of that event in the long-run or infinite sequence, then on what grounds can we justify asserting that a frequency probability has any relevance to any observed sequence? These, and other similar problems, are more generally known as reference class problems.

Of course, Scioscia most likely did take into consideration the different conditions that are causing the reference class problem in this example. Scioscia can take solace in that he was not the first to face this problem. In fact, Karl Popper, among others, recognized the limitations of the frequentist interpretation of probability, and proposed revisions to address these problems. It is worth a brief look at these revisions.

Propensity interpretation of probability

Frequency theorists were (and are) aware of these problems facing frequency interpretations of probability, and have proposed revisions to address these problems. One of the first of these modifications was Popper's *propensity theory* (Popper, 1957; Popper, 1959b). One of the primary motivations driving Popper to develop his propensity theory was the desire to assign 'physically real' probabilities to singular events (Popper, 1959b, p. 28):

> … the interpretation of the two-slit experiment … ultimately led me to the propensity theory: it convinced me that probabilities must be 'physically real' – that they must be physical propensities, abstract relational properties of the physical situation, like Newtonian forces. … Now these propensities turn out to be *propensities to realise singular events*. It is this fact which led me to reconsider the status of singular events within the frequency interpretation of probability.

So the propensity theory of probability, like the frequentist interpretation, is a metaphysical theory.

Rather than identify probability with the long-run relative frequency of an event, propensity theory identifies probability as the propensity of an event to occur under specified conditions (Popper, 1959b, p. 34):

> The frequency interpretation always takes probability as relative to a sequence which is assumed to be given; and it works on the assumption that a probability is *a property of some given sequence*. But with our modifications, the sequence in its turn is defined by its set of *generating conditions*; and in such a way that probability may now be said to be *a property of the generating conditions*.

Popper recognized the radical implication of this with regard to assigning probability to a singular event (Popper, 1959b, p. 34):[11]

> But this makes a great difference, especially to the probability of a singular event (or an 'occurrence'). For now we can say that the singular event a possesses a probability $p(a|b)$ owing to the fact that it is an event produced, or selected, in accordance with the generating conditions b, rather than owing to the fact that it is a member of a sequence b. In this way, a singular event may have a probability even though it may occur only once; for its probability is a property of its generating conditions.

So we can reconsider Siddall & Kluge's dialogue between the probabilist and the historian mentioned above. If we understand the true tree as a singular event, we could understand the probability of that event as a propensity of its generating conditions. Or, in

---

[11] Reading 'a probability $p(a|b)$' as 'a probability p of a given b'.

terms more familiar to a systematist, we might understand the probability of the actual character state distributions as a propensity of the conditions in which this distribution was produced. These conditions are generally described by biologists as models of evolution (or, in the case of Cladists, as the propensity of real lineages to produce traits in a 'parsimonious' pattern). Again, looking to Popper (Popper, 1959b):

> This modification of the frequency interpretation leads almost inevitably to the conjecture that probabilities are dispositional properties of these conditions – that is to say, propensities. This allows us to interpret the probability of a *singular* event as a property of the singular event itself, to be measured by a conjectured *potential or virtual* statistical frequency rather than by an *actual* one.

There are many varieties of propensity theory, and some propensity theorists claim that any revision of a frequency interpretation that accommodates the reference class problem ought to be considered some kind of propensity theory (Gillies, 2000). However, as there are no explicit propensity phylogenetic techniques,[12] discussion of propensity theory will have to be truncated in favor of moving on.

Subjective interpretations of probability

Recall Siddall & Kluge's appeal to an example from baseball (1997, p. 314-15, emphasis added): "More to the point, Alomar either will or he will not get a hit and there is no probability that can be assigned to that one event: *betting on one event alone is foolish*." This is misleading, and confuses the issue at hand. After all, whether a bet is foolish or not depends upon the odds one has been offered. It may be foolish to make most bets as a patron of a casino, but getting ten-to-one odds on a fair coin landing heads might be more reasonable. Bayesians extend this principle to develop an epistemic or (more familiarly) subjective theory

---

[12] Though I do not have room to discuss this point further, the ramifications for the debate at hand are potentially quite significant. Indeed, were statistical phylogeneticists to turn out to be using some kind of propensity theory of probability, then whether they were assigning propensities to trees (phylogenetic or otherwise) becomes a rather banal and less controversial point.

of probability (Gillies, 2000).[13] This is important for at least two reasons. First, it is evident that people do, in fact, make bets on singular events all the time.[14] Secondly, Bayesians claim that an examination of betting behavior demonstrates exactly how subjective probability makes contact with singular events.

For Bayesians, subjective probability is a reflection of the degree of belief a person has in a proposition, and these beliefs can be measured and quantified. Subjective probabilities can be modified by conditionalizing upon evidence. This conditionalization process, it is claimed, can be approximated by using Bayes Theorem.[15] Prior to conditionalization, subjective probabilities are called *prior subjective probabilities*, with the resultant conditionalized beliefs called *posterior subjective probabilities*. One feature of Bayesianism is that a posterior subjective probability becomes the prior subjective probability for the next conditionalization event.

So, if Scioscia were a Bayesian then he *could* assign a subjective probability to the proposition that Bonds will get a hit in a particular at-bat. Just as the propensity theorist will need to take into account as many specific conditions as possible to determine the propensity of a batter to get a hit, the Bayesian will need to conditionalize on as much evidence as is relevant *e.g.*, how Bonds fares at home, away, with runners on base, in the

---

[13] The mechanics of how to determine a betting quotient are often given in the form of a 'Dutch Book' example. We are asked to imagine a situation in which Mr. A must make a bet with Ms. B. For any given event, E, Mr. A must propose a betting quotient q. Ms. B will choose a stake, S, but Mr. A will not know whether the stake S is positive or negative. For any event, E, Mr. A must pay Ms. B qS. If the event E occurs, then Ms. B pays Mr. A S in return. So for any given event E that occurs, Ms. B's gain is qS-S, and Mr. B's gain is S-qS. It can be shown that if Mr. A sets his betting quotient in such a way that it violates the axioms of probability, then he will be subject to a "Dutch book" – i.e., Ms. B will be able to set up the bet in such a way that she will gain no matter the outcome. *E.g.*, if Mr. A sets q at 1.5, then Ms. B should always make S positive, as qS will always be greater than S (and Ms. B's gain will always be positive).

[14] Of course, it might be the case that these people are acting foolish when they do so.

[15] Bayes theorem is $p(h|e) = p(e|h)p(h) / p(e)$, reading '$p(h|e)$ as 'the probability p of h given e.' h is the proposition or hypothesis to which the probability is being assigned, and e is the event or evidence upon which the prior probability is being conditionalized. $p(h)$ is the prior probability of h, whereas $p(h|e)$ is the posterior probability of h, *i.e.*, the probability of hypothesis h conditionalized on evidence e.

playoffs, etc. The amount of relevant information here reflects the complexity and difficulty of assigning subjective probabilities to things like Bonds getting-a-hit in the World Series. But this conforms with experience. It is notoriously difficult to predict when a baseball player will get a hit or not. This, of course, is part of the appeal of baseball (and the source of many baseball debates). The best managers are those who know how to recognize which information is relevant and take it into consideration appropriately. If it were otherwise, baseball would be a much less exciting game – or at least must easier to manage. How much confidence a Bayesian will place in a subjective probability will be relative to how much information is available to them. For both the propensity theorist and the subjectivist, the better acquainted they are with a system the more confident they will be in the probabilities they assign to particular events.

The purpose of this broad overview of interpretations of probability was to evaluate premise P2 of the argument from probability: *Probabilities cannot be assigned to unique historical entities*. As it stands, P2 is not formulated precisely enough to reflect the different interpretations of probability available to the scientist, and needs to be revised. This revision can take the form of the following disambiguation, and on some terms of this disambiguation Siddall & Kluge's argument is no longer valid (though to be fair to Siddall & Kluge, P2a most closely resembles the thesis offered in their paper):

P2a Frequentist probabilities cannot be assigned to singular events.
P2b Propensities can be assigned to singular events.
P2c Bayesian posterior probabilities can be assigned to descriptions of singular events.

4.3. Concerning P3: Therefore, phylogenetic trees are not the kinds of things to which probabilities can be assigned.

The conclusion at P3 no longer follows from the revised premises. Instead we get the following:

P3a Therefore, the true tree is not the kind of thing to which a frequentist probability can be assigned.

This conclusion seems to be consistent with what Siddall & Kluge are claiming, especially in light of their discussion of probabilist versus historical thinking (see above). The object in question, the true tree, is an historic individual, insofar as lineages can be thought of as individuals. Siddall & Kluge are correct in making this point about frequentist probability, and they are right that this would, indeed, be something to be concerned about.

However, as previously described, the problem faced by systematists is epistemic. To reflect this, we need to insert a new premise, something like:

P3b Phylogenetic trees, as hypotheses of the structure of parts of the true tree, are models of ancestral states of current taxa. Such models are more or less well supported by evidence.

P3b, then, formally recognizes the process of evaluating phylogenetic trees as an epistemic problem, *not*, as Siddall & Kluge imply, a metaphysical question. Systematists are working under the assumption that of the possible phylogenetic trees, at least one relevantly captures the structure of the true tree.

Inserting this new premise serves to highlight that the central problem faced by systematists is to evaluate amongst the different possible phylogenetic trees. Typically this is done using parsimony, likelihood or Bayesian cladistic analysis. So now we are in a position to evaluate Siddall & Kluge's fourth premise.

4.4. Concerning P4: Statistical methods claim to assign probabilities to phylogenetic trees.

Recall that Siddall & Kluge's primary concern was that statistical techniques were assigning frequency probabilities to the true tree. In the original argument, this concern was located in premise P4. But look at what has happened. By carefully distinguishing between the true tree and phylogenetic trees, even if we accept premise P4 as it stands Siddall & Kluge's concern is no longer present! That said, premise P4 is too blunt and stands in need of revision in light of recognition of the fact that phylogenetic trees (irregardless of the method by which they were constructed) are hypotheses about the ancestral states of current taxa. A new premise is needed that reflects both (i) that phylogenetic techniques evaluate hypotheses and provide justification for selecting one hypothesis over others (whether that justification is in terms of most corroborated, best supported, etc.); and (ii) that there are, at present, three major techniques found in contemporary systematics (*i.e.*, parsimony, ML, and Bayesian analysis). So, without worrying too much about the details, lets look at how each phylogenetic technique provides justification for selecting one among many possible phylogenetic trees.

Parsimony

Parsimony, as described earlier, is a cladistic technique that selects from among the possible phylogenetic trees that tree which requires the least number of evolutionary steps and is still consistent with the data. The most parsimonious phylogenetic tree, then, is said to be the most corroborated tree, and, thus, an inference to the structure of the true tree is justified on a Popperian falsificationist scientific methodology.

It is worth noting that Cladists may be implicitly appealing to some kind of propensity theory. Parsimony techniques make certain assumptions about the process of evolution. Namely, that lineages have the propensity to produce a pattern of character state distributions that reflect the minimal amount of evolutionary events. This appeal to propensity, though, would be perfectly consistent with Popperian falsificationism (indeed, it would even be expected).

Maximum Likelihood (ML)

Systematists using ML do not assign any kind of probabilities to either the true tree or to phylogenetic trees. ML assigns a likelihood value to phylogenetic trees conditional on the data. Likelihoods are not the same statistical measure as probabilities (Sober, 2000).[16] For one thing, Likelihoods are neither consistent with nor a consequence of the axioms of probability; it is merely an alternative statistical method for evaluating hypotheses. The likelihood of a phylogenetic tree conditional on the observed data, $L(h|e)$, is equal to the probability of the observed data conditional on that phylogenetic tree, $p(e|h)$.[17] Though it might appear that ML advocates are guilty of just what Siddall & Kluge are concerned about – assigning a probability to the actual character state distributions – there are two reasons to think this is not true. The first is that a probability is being assigned only to *observed* character states, *not* to the actual character state distributions *in toto*. Small solace, perhaps, but solace none the less. Popperians, however, ought to take this very seriously. Popper accepted that all observational statements are theory laden and, thus, themselves fallible, and, as such, should be recognized as epistemic and not metaphysical claims (Popper, 1989) (see (Howson and Urbach, 1993, p. 132) for more discussion on this point). The other reason Siddall & Kluge's concerns are not applicable here is that the probability of the observed data is

---

[16] That the likelihood of a hypothesis *h* given evidence *e* is different than the probability of *h* given *e* can be drawn out with a simple example. Presume you were attending a philosophy talk, and knew both (i) that 99% of philosophy talks are given by philosophers, and (ii) that 99% of philosophers read their papers verbatim at philosophy talks. Suppose that at this particular philosophy talk, the paper was *not* being read verbatim and you wanted to know whether the speaker was a philosopher. We can set up the problem as follows: h = that the presenter is a philosopher; e = the paper is not being read verbatim; $p(h) = 0.99$; $p(e) = 0.01$. The likelihood that the speaker is a philosopher, $L(h|e)$, is equal to the probability of the evidence given the hypothesis, $p(e|h)$, which equals 0.01. So there is a very low *likelihood* that the speaker is a philosopher. The probability that the speaker is a philosopher given that the paper is not being read verbatim, $p(h|e)$, can be determined using Bayes theorem. $p(h|e) = p(e|h)p(h) / p(e) = (0.01)(0.99) / (0.01) = 0.99$. So the *probability* that the speaker is a philosopher is very high.

[17] The interpretation of probability used here, and elsewhere in ML techniques, may vary. In what follows, I will simply use the generic term 'probability' to refer to this range of possible interpretations of probability. It should be apparent by now that which interpretation of probability is used will have great consequences regarding whether Siddall & Kluge's concerns are applicable.

ranging over a set of possible character state distributions – just the criteria Siddall & Kluge demand for an assignment of frequency probability (1997, p. 317):

> For frequency probability to apply to phylogeny there has to be a set of *simultaneously* possible trees….

To see why this is so, a deeper look at how likelihoods are assigned to trees is needed.

In ML, a model of evolution must be specified. Part of the specification of this model includes the probability of changing from one character state to another on any given branch of a phylogenetic tree (*e.g.*, changing from one nucleotide state to another). Recall that a phylogenetic tree is an inference of ancestral states. So, for any hypothetical ancestral conditions (*i.e.*, any phylogenetic tree), the probability of any distribution of character states from those conditions can be determined. This is what is measured by p(e|h). Recall also that the actual state of affairs is typically understood as one among many possible state of affairs. To determine the likelihood of a particular phylogenetic tree given a particular set of data (actual or otherwise), is to ask what the probability of that data set is given that phylogenetic tree. The probability is ranging over a set of possible distributions of data conditionalized on phylogenetic trees and a model of evolution (Swofford et al., 1996). The phylogenetic tree with the highest likelihood value conditional on the data is then claimed to be the best supported hypothesis of the structure of the true tree, and is called the ML phylogenetic tree.

Bayesian Phylogenetic Analysis

As the name would imply, Bayesian phylogenetic analysis evaluates the possible phylogenetic trees using posterior probabilities. The range of possible phylogenetic trees makes up the parameter space over which the posterior probabilities are distributed. Bayesian phylogeneticists use Markov Chain Monte Carlo (MCMC) algorithms to approximate the distribution of posterior probabilities over the parameter space (Larget and Simon, 1999;

Huelsenbeck et al., 2002). The phylogenetic tree (or consensus trees) with the highest posterior probability is then said to be the best supported hypothesis of ancestral states conditional on the data.

The above descriptions show how the different phylogenetic techniques evaluate the possible phylogenetic trees and also justify the inferential claims involved in selecting among the phylogenetic trees. Premise P4 can now be revised to reflect this:

P4a Cladists use parsimony (in a Popperian falsificationist framework[18]) to justify inferential claims about the state of the true tree.
P4b Statistical phylogeneticists use either (i) likelihood techniques utilizing statistical approaches or (ii) Bayesian subjective probability to justify inferential claims about the ancestral conditions of current taxa.

5. Conclusion

Siddall & Kluge's argument from probability fails to tell against statistical techniques in phylogenetics. To recap, Siddall & Kluge's argument from probability runs as follows:

P1 Phylogenetic trees are unique historical entities.
P2 Probabilities can not be assigned to unique historical entities.
P3 Therefore, phylogenetic trees are not the kinds of things to which probabilities can be assigned.
P4 Statistical methods claim to assign probabilities to phylogenetic trees.
P5 Therefore, statistical methods are not applicable to the building of phylogenetic trees.

I have shown that each of these premises is in need of revision or clarification. Upon revision, the argument from probability now runs as follows:

P1a The true tree (*i.e.*, the actual historical lineage of life) is a unique historical entity which is unobservable and, in practice, unknowable.
P1b Phylogenetic trees are hypotheses about the structure of the true tree.
P2a Frequentist probabilities can not be assigned to singular events.
P2b Propensities can be assigned to singular events.
P2c Bayesian posterior probabilities can be assigned to descriptions of singular events.
P3a Therefore, the true tree is not the kind of thing to which a frequentist probability can be assigned.

---

[18] Accepting here as unproblematic Siddall & Kluge's claim of a falsificationist conceptual underpinning.

P3b Phylogenetic trees, as hypotheses of the structure of parts of the true tree, are models of ancestral states of current taxa. Such models are more or less well supported by evidence.

P4a Cladists use parsimony (in a Popperian falsificationist framework) to justify inferential claims about the state of the true tree.

P4b Statistical phylogeneticists use either (i) likelihood techniques utilizing statistical approaches or (ii) Bayesian subjective probability to justify inferential claims about the ancestral conditions of current taxa.

In the revised form of the argument, the conclusion that statistical methods are not applicable to the building of phylogenetic trees does not follow from the premises. The problems that Siddall & Kluge identify with statistical phylogenetics go away upon recognition of the important distinctions between (i) the true tree and phylogenetic trees, and (ii) epistemic and metaphysical possibility. Parsimony and statistical techniques are competing methods for approaching the epistemic problem of evaluating amongst possible phylogenetic trees, *i.e.*, hypotheses of the true tree. None of these methods, however, are inappropriately applying probabilistic thinking to phylogenetic problems.

A final point is worth noting. I have downplayed the potential points of conflict within statistical phylogenetics. The recent introduction and development of Bayesian phylogenetic techniques presents a challenge to ML advocates. How the adoption of these Bayesian techniques plays out in theoretical systematics bears watching. I anticipate that debates surrounding this issue will grow more central as Bayesian techniques are more widely adopted, possibly supplanting the debate between Cladists and statistical phylogenetics.

Works Cited

Baum, D. A. 1998. Individuality and the Existence of Species through Time. *Systematic Biology* 47:641-653.

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. *Systematic Biology* 50:305-321.

Eldredge, N., and J. Cracraft. 1980. *Phylogenetic Patterns and the Evolutionary Process : Method and Theory in Comparative Biology.* Columbia University Press, New York.

Farris, J. S. 1983. The Logical Basis of Phylogenetic Analysis. Pages 7-36 *in* Advances in Cladistics, Volume 2: Proceedings of the Second Meeting of the Willi Hennig Society (N. I. Platnick, and V. A. Funk, eds.). Columbia Univ. Press, New York.

Felsenstein, J. 1978. Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading. *Systematic Zoology* 27:401-410.

Felsenstein, J. 1981. Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *Journal of Molecular Evolution* 17:368-376.

Felsenstein, J. 2001. The Troubled Growth of Statistical Phylogenetics. *Systematic Biology* 50:465-467.

Ghiselin, M. T. 1974. A Radical Solution to the Species Problem. *Systematic Zoology* 23:536-44.

Gillies, D. 1990. Bayesianism Versus Falsificationism. *Ratio (New Series)* III:82-98.

Gillies, D. 2000. *Philosophical Theories of Probability.* Routledge, London ; New York.

Grunbaum, A. 1976. Is the Method of Bold Conjectures and Attempted Refutations *Justifiably* the Method of Science. *The British Journal for the Philosophy of Science* 27:105-36.

Hennig, W. 1979. *Phylogenetic Systematics*. University of Illinois Press, Urbana.

Hillis, D. M., J. P. Huelsenbeck, and C. W. Cunningham. 1994. Application and Accuracy of Molecular Phylogenies. *Science (Washington D C)* 264:671-677.

Howson, C., and P. Urbach. 1993. *Scientific Reasoning : The Bayesian Approach*, 2nd edition. Open Court, Chicago.

Huelsenbeck, J. P. 1997. Is the Felsenstein Zone a Fly Trap? *Systematic Biology* 46:69-74.

Huelsenbeck, J. P., and D. M. Hillis. 1993. Success of Phylogenetic Methods in the Four-Taxon Case. *Sys. Biol.* 42:247-264.

Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. 2002. Potential Applications and Pitfalls of Bayesian Inference of Phylogeny. *Systematic Biology* 51:673-688.

Huelsenbeck, J. P., and B. Rannala. 1997. Phylogenetic Methods Come of Age: Testing Hypotheses in an Evolutionary Context. *Science (Washington D C)* 276:227-232.

Hull, D. L. 1976. Are Species Really Individuals? *Systematic Zoology* 25:174-91.

Hull, D. L. 1983. Karl Popper and Plato's Metaphor. Pages 177-190 *in* Advances in Cladistics, Volume 2: Proceedings of the Second Meeting of the Willi Hennig Society (N. I. Platnick, and V. A. Funk, eds.). Columbia Univ. Press, New York.

Hull, D. L. 1988. *Science as a Process : An Evolutionary Account of the Social and Conceptual Development of Science.* University of Chicago Press, Chicago.

Jubien, M. 1997. *Contemporary Metaphysics : An Introduction.* Blackwell Publishers, Malden, Mass.

Jubien, M. 2001. Thinking About Things. *Onus (Supplement)* 15:1-15.

Kitcher, P. 1982. *Abusing Science : The Case against Creationism.* MIT Press, Cambridge, Mass.

Kluge, A. G. 1997a. Sophisticated Falsification and Research Cycles: Consequences for Differential Character Weighting in Phylogenetic Systematics. *Zoologica Scripta* 26:349-360.

Kluge, A. G. 1997b. Testability and the Refutation and Corroboration of Cladistic Hypotheses. *Cladistics* 13:81-96.

Kluge, A. G. 2001. Philosophical Conjectures and Their Refutation. *Systematic Biology* 50:322-330.

Kripke, S. A. 1980. *Naming and Necessity*. Harvard University Press, Cambridge, Mass.

Kuhn, T. S. 1970. Logic of Discovery or Psychology of Research? Pages 1-23 *in* Criticism and the Growth of Knowledge (I. Lakatos, and A. Musgrave, eds.). Cambridge University Press, Cambridge.

Kuhn, T. S. 1996. *The Structure of Scientific Revolutions*, 3rd edition. University of Chicago Press, Chicago, IL.

Lakatos, I. 1970. 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 University Press, Cambridge.

Larget, B., and D. Simon. 1999. Markov Chain Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees. *Mol. Biol. Evol.* 16:750-759.

Lewis, P. O., and D. L. Swofford. 2001. Back to the Future: Bayesian Inference Arrives in Phylogenetics. *Trends in Ecology & Evolution* 16:600-601.

Popper, K. R. Year. The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory. *in* Observation and interpretation; a symposium of philosophers and physicists. Butterworths Scientific Publications, University of Bristol:65-70.

Popper, K. R. 1959a. *The Logic of Scientific Discovery*. Basic Books, New York,.

Popper, K. R. 1959b. The Propensity Interpretation of Probability. *The British Journal for the Philosophy of Science* 10:25-42.

Popper, K. R. 1989. *Conjectures and Refutations*, 5 edition. Routledge.

Salmon, W. C. 1998. Rational Prediction. Pages 433-444 *in* Philosophy of Science: The Central Issues (M. Curd, and J. A. Cover, eds.). W. W. Norton & Company, Inc., New York.

Siddall, M. E., and A. G. Kluge. 1997. Probabilism and Phylogenetic Inference. *Cladistics* 13:313-336.

Sober, E. 1983. Parsimony Methods in Systematics. Pages 37-48 *in* Advances in Cladistics, Volume 2: Proceedings of the Second Meeting of the Willi Hennig Society (N. I. Platnick, and V. A. Funk, eds.). Columbia Univ. Press, New York.

Sober, E. 1988. *Reconstructing the Past : Parsimony, Evolution, and Inference.* MIT Press, Cambridge, Mass.

Sober, E. 2000. *Philosophy of Biology*, 2 edition. Westview Press, Boulder, CO.

Sokal, R. R., and P. H. A. Sneath. 1963. *Principles of Numerical Taxonomy*. W. H. Freeman, San Francisco,.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Chapter 11: Phylogenetic Inference. Pages 407-514 *in* Molecular Systematics (D. M. Hillis, C. Moritz, and B. Mable, eds.). Sinauer, Sunderland, MA.

Urbach, P. 1991. Bayesian Methodology: Some Criticisms Answered. *Ratio (New Series)* IV:170-184.