Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 35 (2005) 271–280
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Review

Falsifications and corroborations: Karl Popper's influence on systematics

Kevin G. Helfenbein *, Rob DeSalle

American Museum of Natural History, Department of Invertebrate Zoology, Central Park West at 79th Street, New York, NY 10024, USA

Received 21 December 2004

Abstract

Over the last three decades, the philosophy of Karl Raimund Popper has had a strong influence on the field of systematic biology. Unequivocally, no other philosopher's work has had such an influence during this formative period in systematics. Much, but not all, of the early discourse on Popper and systematics dealt with the philosophical basis of systematics as a science. More recently Popper's work has been discussed in the systematics literature in relation to specific methodologies such as parsimony and maximum likelihood. In this paper, we provide the reader with a concise summary of Popper's ideas relevant to systematics, review the systematic literature invoking or declining Popper's importance to the field, and make a recommendation for the future course of philosophical thinking in systematics. We try to make clear various authors' interpretations of Popper's work and how those interpretations have impacted systematic thought. Although the reader may come away from this review with a clearer idea of Popper's relevance or lack thereof, our primary hope is that the reader will be compelled to question him- or herself about the philosophical basis of the systematic work that he or she does, and to delve into the literature herein cited. We begin by presenting a synopsis of Popper's philosophical views to allow those views to be placed in the context of systematics.

Keywords: Karl Popper; Falsification; Corroboration; Systematics; Philosophy of science

1. Popper and philosophy

The air was full of revolutionary slogans and ideas, and new and often wild theories.

Popper (1965)

1.1. Popper's philosophical origins

The time and place of Popper's birth (Vienna, 1902) may in some part be responsible for Popper's broad interests in music, politics, philosophy, and science. He was interested in many scientific and sociological theories proposed at the time. The theories that most interested him were Adler and Freud's psychological theories, Marx's theory of history, and Einstein's theory of relativity. Popper's thoughts on these theories, particularly their claims as *scientific* theories, led him to formulate, arguably, the most important philosophy of science in the twentieth century. Popper viewed one of the above four theories as different than the others, and did so in the context of an older philosophical problem, Hume's Problem of Induction [David Hume 1711–1776, Scottish empiricist philosopher]. Induction is commonly defined as logic or thought proceeding from the specific to the general, i.e., forming any generalized expression (hypothesis, theory, conclusion, etc.) after making a series of observations. Alternatively, induction can be thought of as confirmation or verification of a general statement through observational iterations. For example, the repeated observations of white swans

* Corresponding author. Fax: +1 212 769 5277.
E-mail address: kgh@amnh.org (K.G. Helfenbein).

1055-7903/$ - see front matter
doi:10.1016/j.ympev.2005.01.003

and white swans only confirms the generalized statement "all swans are white." But does it? Hume questioned whether one could reasonably justify that the universal set described in a generalization (here that all swans are white) could be verified from the observation of a finite number of the set. Popper saw the "problem" as no problem at all, but as a damnation of induction as a mode of logical inference.

Among the four theories that interested Popper in his early years, he saw only one that was not inherently inductive, namely Einstein's Theory of Relativity. What set Einstein's theory apart from Marx's, Freud's, and Adler's was that it could be tested in such a way that it could, in principle, be shown to be false. Thus, Popper made the claim that falsifiability was the point of demarcation between science and pseudoscience, and that all theories that claimed to be scientific must be falsifiable (though as we shall see below falsifiability and non-verifiability apply only to some theories). Furthermore, Popper claimed an inverse relationship exists between falsifiability and probability; thus, the most improbable theories are the most falsifiable. These theories, ones that Popper referred to as bold or wild conjectures, are the ones to be preferred among scientists.

1.2. Corroboration

Popper's preference of falsifiability over verifiability may evoke the hypothetico-deductive method. However, two points separate Popper's deductive falsificationism and hypothetico-deductivism: as mentioned above Popper prefers improbable theories whereas the hypothetico-deductivist prefers more probable hypotheses. In addition, the hypothetico-deductivist considers unfalsified hypotheses to be confirmed theories. The latter postulate is unacceptable to Popper (Salmon, 1967). Yet, surely Popper must have seen scientific methodology as something more than the ability to falsify hypotheses. And indeed, he did.

Hypotheses, according to Popper, should stand up to the most severe tests. In the cases where hypotheses were not falsified under those circumstances, Popper introduced the term *Grad der Bewährung*. Initially translated by Carnap [Rudolph Carnap 1891–1970 German logical positivist] as "degree of confirmation," Popper rejected this wording and in its place used "degree of corroboration" (Popper, 2002, p. 248). Understanding what he meant by "corroboration," then, is paramount to understanding the full scope of Popper's philosophy of science.

The degree to which a hypothesis stands up to severe tests, the appraisal of the worth of the hypothesis, is its degree of corroboration. Hypotheses, other than tautologies, that have been tested and not falsified have been corroborated. The standing of such a hypothesis, though, is not one of a confirmed hypothesis. Specifically it should not be considered a "true" statement following the critical testing. Truth, according to Popper (2002), is atemporal—what is true now was true in the past and will be true in the future. Corroboration is temporal: one test may have corroborated a hypothesis yesterday, another more severe test corroborated it today, and a most severe test may corroborate the hypothesis tomorrow. Thus, the rejection of the word "confirmation" is in perfect agreement with Popper's attitude toward hypothesis testing and his skeptical attitude toward our ability to claim discovery of 'truth.'

Popper went further than this qualitative description of corroboration, providing a formulaic definition. Before introducing the formula, four terms need to be examined: probability (p), background knowledge (b), empirical evidence (e), and hypothesis (h). Popper distinguishes between numerical probability and logical probability. The former refers to frequency considerations such as the probability of rolling a particular number on a true die. The latter, logical probability, Popper defined thusly: *The logical probability of a statement is complementary to its degree of falsifiability*: it increases with decreasing degree of falsifiability (Popper, 2002, p. 102, his italics). Popper (1983) considered b as any relevant knowledge accepted even provisionally when testing a hypothesis (i.e., b itself is not currently being tested). e should not be too probable given b alone, i.e., the data should be unexpected given what you already know. The best h as discussed above, should be a bold conjecture, a "hopeful monster" of potential explanation, rather than one that offers little that is not already known.

Given these factors, consider the following (Popper, 1983):

$$C(h,e,b) = \frac{p(e,hb) - p(e,b)}{p(e,hb) - p(eh,b) + p(e,b)},$$

where, for example, '$p(e,hb)$' is read "the probability of the evidence given the hypothesis and the background knowledge." Of the denominator, Popper (1983, p. 240) says, "... the denominator has no [simple intuitive] significance ... it seems to be the simplest normalization factor ...." Consider only the numerator then: $p(e,hb) - p(e,b)$, the probability of the evidence given the hypothesis and background knowledge minus the probability of the evidence given the background knowledge alone. As the probability of e becomes greater due to h relative to b, the value obtained is positive (corroboration). If the probability of e in light of h is less that that given b alone, a negative value is obtained (falsification). If both probabilities are the same, the value is zero (tautology). Popper stated that "the support given by e to h becomes significant only when $p(e,hb) - p(e,b) \gg 1/2$." (Popper, 1983, p. 240) (see Fig. 1).

This value, so far as we can determine, is arbitrary. It clearly reflects Popper's feeling that the hypothesis should be grand. This notion aside, it would seem that

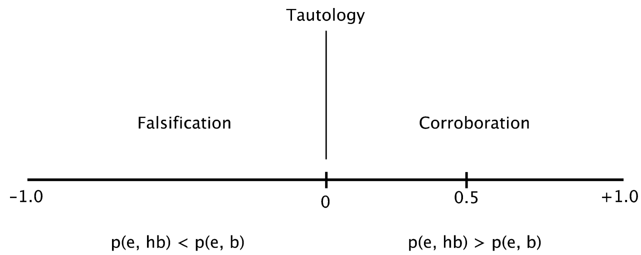

Fig. 1. Schematic of the outcomes of Popper's corroboration formula. Popper considered 0.5 a significant value in this scheme.

any case where $p(e, hb) > p(e, b)$ is positive would yield a corroborating value. In fact, it may seem intuitive that science does often proceed in just this way, in small steps rather than Popperian leaps.

1.3. Popper and real science

This brings up an important point: Popper's work is not a philosophical analysis of science as it is practiced, but rather a prescription of how the logic of science should proceed. From his viewpoint, all science is a matter of conjecture rather than application of knowledge per se, theory as an end for itself (Putnam, 1974). Or, as Popper (2002, p. 37) himself states, "The empirical sciences are systems of theories. The logic of scientific knowledge can therefore be described as a theory of theories."

That Popper ignores the practical aspects of scientific research, an endeavor in which practice is primary (Putnam, 1974), is just one critical comment directed toward Popper and his philosophy. A complete survey of criticisms of Popper is beyond the scope of this review. However, for a sense of the breadth of the criticisms of Popper's philosophy, consider the following: falsification requires prediction from theory, but prediction does not always follow from a theory, rendering falsification impossible in those cases (Putnam, 1974). Popper's system of falsification/corroboration relies on critical tests of individual statements. Scientific research programs are composed of systems of interacting theories such that critical tests do not exist—non-corroboration is not necessarily falsification of a particular statement (Lakatos, 1978). Scientists who do not try to minimize the probability of hypotheses do not do so exclusively, but also seek content in non-minimally probable statements (Franklin, 2001). Popper's corroboration is a form of non-demonstrative inference; ironically, in justifying it, he justifies induction, another form of the same (Salmon, 1967).

Given these and other critiques of Popper (see, for example, Sober's (1988), footnote p. 121), and given the fact that few, if any, scientific research programs have progressed by molding a particular scientific field of inquiry to one particular philosophy, it would seem unlikely that any scientist would suggest that research should proceed (strictly) according to Popper's views. Yet this has been suggested. Herein, we will review the literature concerning Karl R. Popper's powerful influence on the field of biological systematics that commenced when most systematic studies utilized morphological data, and has continued, and in fact intensified, into these contemporary times dominated by molecular phylogenetic studies.

2. Popper and systematics

... as Thomas Kuhn has argued, when scientists fall out they frequently resort to philosophy, both to justify their own positions and to attack their opponents.

Ruse (1979)

2.1. Early appeals to Popper

Prior to the 1970s, Popper's work did not directly impact systematic thought and methodology. The state of systematics in this period has been characterized as inductionist/verificationist (Bock, 1973). Popper entered systematics when Bock argued that classical evolutionary classification should prevail as a classification system over phenetics and cladistics due to its (or so Bock stated) consistency with Popper's ideas (Bock, 1973), and, specifically, its ability to combine degree of similarity and phylogenetic sequence of events. Although Bock did not make a compelling case that this is so ("Space does not permit a further elaboration of details, but a general analysis of the theory of evolutionary classification convinces me that it is consistent with Popper's basic ideas."), he raised issues that are still important to the philosophy of systematics, and helped begin a quest for a logical framework for modern systematics (see also Cracraft, 1978).

First and foremost of the questions Bock raised is what are the severe, falsifying tests of hypotheses of classifications, phylogenies, and homology? Second, he highlights Popper's statement (see Popper, 1965) equating a theory's higher degree of empirical content with its testability and states that "The scientific theory [that meets the criterion of high degree of empirical content] as the foundation for all biological comparison is the Darwinian theory of evolution" (Bock, 1973).

Regarding the first point, Bock emphasizes that homology is the underlying principle of comparative biology, and that the recognizing criterion of homology is similarity. Wiley (1975) suggests that Bock equates the recognition of homology to its corroboration/falsification (as Bock offers no other clear-cut solution to the problem) by a similarity criterion. He points out that similarity is not the most severe test for homology;

rather, the most severe test of a hypothesis of homology is other hypotheses of homology (Wiley, 1975). This is the first clear statement of Popperian testability/falsifiability in systematics.

Popper's philosophy underlies Gaffney's (1975) phylogeny of higher categories of turtles, one of the earliest, and perhaps the first such study, exhibiting the philosopher's influence. Gaffney used a cladistic approach, stating that it seemed most "compatible" with Popperian testability. Interestingly, he considered a discussion of the cladistic methodology used in the study necessary "... so that other workers will be able to test the hypotheses presented using different methods" (Gaffney, 1975). Since that study, few systematists who take a philosophical approach to their work have advocated the use of multiple phylogenetic methodologies to test hypotheses.

2.2. Is systematics Popperian?

These attempts at defining systematics in Popperian terms begged an important question: is systematics the type of scientific inquiry Popper had in mind when formulating his philosophy of science (Kitts, 1977)? What exactly are the types of theories or statements that can only be falsified, not verified? Popper makes it clear that only strictly universal statements have this property and that "Scientific theories are [strictly] universal statements" (Popper, 2002, p. 37). Universal statements refer or apply to an unlimited number of individuals in space and time. For example, Newton's laws of motion apply to all planets that have ever existed or will exist. Numerically universal statements, on the other hand, are conjunctions of singular statements, statements that refer only to certain finite regions of space and time (Popper, 2002). Popper notes that it is his convention that scientific laws are strictly universal. His decision on this point is determined by the fact that only strictly universal statements, not numerically universal statements, are non-verifiable and thus strictly only falsifiable.

Popper (2002, pp. 42–43) states, "It is usual to elucidate [the distinction between universal and singular statements] with the help of examples of the following kind: 'dictator,' 'planet,' 'H_2O' are universal concepts ... 'Napoleon', 'the earth,' 'the Atlantic' are singular or individual concepts ... ." Though recent discussions regarding the nature of species have been contentious, including at least one call to eliminate this taxonomic rank (Mishler, 1999), traditional views on the subject that have informed the arguments regarding Popper and systematics suggest that to Popper's first list we could add 'species' and to the second any specific example, say '*Homo sapiens*.' Clearly, individual species, the fundamental units of systematics and classification systems, are individuals. But what of systematic hypotheses and classifications themselves? Kitts (1977) argued that these too are not strictly universal: "whatever it is that classifications have to say on the subject of [the association of character states], it is not without temporal restriction." Because classifications are conjunctions of singular statements, they are in principle both verifiable and falsifiable (Kitts, 1977). So too phylogenies that "like any history ... purports to be an account of the spatial, temporal, and causal relationships among events" (Kitts, 1977). Kitts's thesis rings a cautionary note. As he puts it (Kitts, 1977): "... if taxonomists mistakenly come to see their task as the formulation of strictly universal statements to be adduced in predictions and explanations, it may well have a significant and detrimental effect upon systematic[s]." Others have argued to the contrary.

In a review of *The Logic of Scientific Discovery*, Platnick and Gaffney (1977) argue that although phylogenetic hypotheses (trees) may be singular statements, cladistic hypotheses (cladograms) are universal. They reasoned that, rather than making a claim about a particular historical relationship, a cladogram makes a universal statement "about the relative degrees of relationship among taxa that, if true, will be true whenever and wherever in the universe members of those taxa occur"; this viewpoint, however, confounds the commonly held idea of the historical with respect to "historical relationship." Among a set of direct replies to Kitts (Cracraft, 1978; Nelson, 1978; Patterson, 1978), Cracraft pointed out that Popper himself (in Popper, 2002) admitted an ambiguity of the distinction between strictly universal and numerically universal statements: if a statement is verifiable in principle only but not in practice, might we not be able to consider it strictly universal? Cracraft and Nelson perceive cladograms as hypotheses that could not possibly be verified, thus putting them within the Popperian framework. This would seem to be bending the rules slightly—the inability to verify has more to do with the historical nature of the hypotheses rather than their universality—but illustrates the point that it is not necessary to be strictly Popperian to start injecting a logically sound framework into one's science. Patterson (1978), claiming that taxa are neither universals nor numerical universals but rather individuals, argued that hypotheses of relationship are not falsifiable, except only in the most trivial sense.

Despite his perception of the inability to falsify systematic hypotheses, Patterson nonetheless advocated in the most general sense a Popperian approach, what has been called the Popperian "spirit" (Settle, 1979), "... striv[ing] to express ... conjectures (of homology, of monophyletic groups and their parts) as clearly and precisely as possible, so that they are accessible to criticism" (Patterson, 1978). The methodology that would leave systematic knowledge-claims most open to criticism, he suggested, is cladistics (Patterson, 1978). Settle's (1979) characterization of the Popperian "spirit" includes not only openness to criticism for the improvement of knowledge, but also the realistic (in the philosophical sense) pursuit of explanation. These qualities are also found in metaphysics (Settle, 1979): that systematics may lie along the border of science and metaphysics is not a problem—Popper (1974b, p. 981) himself states, "... the transition between [the two] is not a sharp one: what was a metaphysical idea yesterday can become a testable scientific theory tomorrow; and this happens frequently ... ."

Hull (1980) continued the discussion of universal versus numerically universal statements, pointing out "The very statement of the problem of induction depends on the difference between genuine and numerical universals .... If laws of nature can refer to numerical universals, then there is no problem of induction." He extends the discussion of distinguishing terms to "natural kinds," a term that appears in a law of nature (a universal statement), and asks whether sister group relationships are universal statements, numerical universal statements, or particular statements (e.g., "Some A are B") and whether taxa are natural kinds. The answers to these questions can, in principle, vary depending on whether or not an investigator stays with the accepted definitions of the terms in question, or whether the words are redefined to suit the investigator. In the latter case, appeals to the philosophical literature can no longer be made—the scientist must take a philosophy-independent path to support scientific claims. Thus, species, characters, monophyletic groups—things that evolve—can only be natural kinds and universal theories if those terms are redefined (Hull, 1980). In doing so, "falsifiability" would be an extension of Popper's notion of the term (Hull, 1980). Perhaps Patterson's criticism of the use of falsifiability in systematic studies could then be sidestepped, but so too could the Popperian worldview.

The Popperian framework of falsification and corroboration is a narrow one. One suggestion to broaden the philosophical framework of systematics came from Ruse (1979) who suggested that consilience also be included. A consilient scientific theory would explain phenomena from diverse areas; it would unify a variety of observations under a single hypothesis (Ruse, 1979). In suggesting this, Ruse implies that it is the relationship between systematics and modern evolutionary theory that would give the former consciliatory power (Ruse, 1979). This opens the door to an important area of discussion—Popper's evaluation of Darwinian evolutionary theory itself.

2.3. Popper on Darwin and evolution

Though there has been much debate regarding the nature of the relationship between Darwinian evolutionary theory and systematics, particularly as to whether one justifies the other (see, for example, Brower, 2000), there is no denying a relationship. Thus, since Popper never spoke directly about systematics, we must look at Popper's writing on evolutionary theory to approximate what his thoughts might have been regarding systematics.

Popper's view on whether or not Darwinian (and neo-Darwinian) evolutionary theory is a scientific or metaphysical hypothesis—or some other class of hypothesis—clearly changed over time (see Stamos, 1996). Popper's primary concern in his first and preeminent work, *The Logic of Scientific Discovery*, is physics. (And indeed, what better discipline is there from which to derive a framework of universal falsifiables?) In fact, the editor's forward to the *Postscript to the Logic of Scientific Discovery* describes that work as "... the culmination of Sir Karl's work in the philosophy of *physics* ... (our italics)" (Bartley, 1983, p. xii). That he was preoccupied with physics may have itself caused Popper problems when he turned to biology, though it may also be true that problems arose because biology is inherently different in some way from physics (a subject we will not go into any further here).

Popper vacillated on evolution not only over the course of years, but within one work, *The Poverty of Historicism*. Initially referring to Darwinian theory as a "brilliant scientific hypothesis" (Popper, 1964, p. 106), Popper then applies his philosophy and notes that it is not a strictly universal statement, nor can one derive testable predictions from it (Popper, 1964). Later, in *Objective Knowledge: An Evolutionary Approach* (Popper, 1979), Popper drew a line of demarcation between Darwinian evolution, a theory historical in nature, and (the superior) Newtonian theory by noting that "Newton formulated a set of universal laws" whereas "Darwin's theory of evolution proposed no such laws" (Popper, 1979, p. 267). Popper characterized Darwinian evolution as untestable, "a metaphysical research program" (Popper, 1974a, p. 134). Unlike the logical positivists of the Vienna Circle with whom he was often associated, Popper considered metaphysics meaningful; thus, he still considered Darwinian theory "invaluable," citing the explanatory power of natural selection in the case of bacterial adaptation to penicillin (Popper, 1974a, p. 137).

Popper turned about face when in 1977 he stated "I have changed my mind about the testability and the logical status of the theory of natural selection ..." (quoted in Stamos, 1996). His change of mind led to a declaration that evolutionary biology was a research program, if not a full-fledged scientific theory (see Stamos, 1996 for a discussion of the implication of this change of mind relative to the Popperian worldview of *The Logic of Scientific Discovery*). Finally, Popper (1980) published a letter in *New Scientist* declaring all historical sciences were testable and thus scientific (see Rieppel, 2003 for a detailed discussion of this letter). Testability, he stated could rely on retrodictions as well as predictions. His letter may have been meant to distance himself from creationists who had used his writings to support their cause (Stamos, 1996). It is important to remember that Popper's criticism of evolutionary biology arises directly from a philosophy of scientific discovery steeped in physics. His analysis of evolutionary biology, one that might apply to many biological subfields including systematics, is notably ill informed (Ruse, 1977) unlike his analyses of problems in physics.

Much of the discourse regarding Popper and systematics in the late twentieth century dealt with general issues of systematics as a science. Although enthusiasm for these arguments had died down somewhat by the early 1990s, Popper and his philosophy remained in systematics' future. The rise of the use of molecular data spurred the development of a new methodology, maximum likelihood (ML). The use of Popper's ideas in philosophical argumentation within systematics was revitalized then by pointed attacks directed at this new methodology.

3. Popper progresses

Whether it is that cladists 'hide behind' philosophy or that likelihoodists 'hide behind' naked operationalism depends on whether or not one needs a rational basis for one's science.

Siddall and Kluge (1997).

3.1. Popper and probabilistic phylogenetic methods: Maximum likelihood

ML approximates the probability of the data (e), typically nucleotide or amino acid sequence, given a phylogenetic tree (h) and a model of evolution (b). Cladistic parsimony, on the other hand, analyzes a data matrix to find nested hierarchies of characters based on shared derived character states. Little or no background knowledge is necessary to perform the analysis. The tree (or trees), i.e., phylogenetic hypothesis that includes the fewest character state changes is preferred over all others. Thus, ML and cladistic parsimony share in common h and, in cases using molecular sequence data, e, but differ in b and in the test performed using these elements.

With respect to Popper, Siddall and Kluge (1997) labeled ML a verificationist approach to phylogenetics. They claimed that because ML assigns a non-zero probability to each tree (hypothesis), it denies, i.e., falsifies, nothing, and lacks explanatory capability. Siddall and Kluge (1997) criticize all frequency probabilistic approaches to solving historical problems, posing, for example, the question "How is 'statistical error' to be interpreted in an evolutionary or other singular framework?" Underlying this question is the view that "history is particular and cannot be described in terms of universal statements about abstract generalities, the task of the historical sciences being one of explanation, not prediction."[1]

In their view, cladistic parsimony, based in Popperian logical probabilism, denies frequency probabilism. They distinguish the two forms of logic by giving an example of each: Bayes's theorem (frequency probabilism), $p(h, e) = p(e, h) \times p(h)/p(e)$, and Popper's degree of corroboration (logical probabilism), in simple form $C(h, e, b) = p(e, h, b) - p(e, b)$. The key difference between the two formulae is the Bayesian posterior probability statement '$p(h, e)$,' the probability of a hypothesis given the evidence.[2] The search for the hypothesis with the highest probability given some evidence is equated by Popper to be "the mistaken solution to the problem of induction." Siddall and Kluge otherwise view likelihoodists' inductive search for 'the correct tree' as a lost cause because of Popper's assertion that one can never know the 'truth,' in this case the true history of a set of phylogenetic relationships.

3.2. Maximum likelihood as Popperian science

De Queiroz and Poe (2001, 2003) countered this attack by not only defending ML as a Popperian enterprise, but also by calling for a reevaluation of cladistic parsimony before allowing it to claim status as a Popperian science. This was the first instance of the defense of ML using Popper's philosophy. There are two key points in De Quiroz and Poe's defense of ML. First, they point out that "... the mistaken solution to the problem of induction involves assigning probabilities to hypotheses, but likelihood does not assign probabilities to hypotheses. Likelihood is not the probability of the hypothesis given the evidence but the probability of the evidence given the hypothesis." In other words, likelihood calculates $p(e, h)$, or, considering ML's use of models, $p(e, hb)$—the first term of Popper's corroboration formula. (Kluge (2001) argues against their interpretation of $p(e, hb)$, explaining that Popper had derived C from absolute (logical) probability, not frequency probability.)

[1] From this description, one might incorrectly infer that the authors argue against, rather than for, applying Popperian philosophy to systematics given the discussions reviewed above.

[2] But note Felsenstein's (2003) comment "... Popper's formula [C] assumes a Bayesian inference framework .... As Popper was an opponent of Bayesianism (Elliot Sober personal communication) his corroboration formula seems fundamentally at odds with his other views." Felsenstein apparently bases his argument on the presence of $p(e, h)$ in both formulae. This term is proportional to Fisher's likelihood (L) (Edwards, 1992), though since it is generally used interchangeably with L in the literature reviewed herein—as though the two were equal rather than proportional—we will use it in the same fashion.

Table 1
Comparison of Popper's corroboration formula and Bayes's formula

Term	Explanation
Popper's corroboration formula [$C = p(e, hb) - p(e, b)$]	
C	Corroboration
$p(e, hb)$	Probability of the evidence given a hypothesis and some background knowledge
$p(e, b)$	Probability of the evidence given the background knowledge alone
Popper explained that a hypothesis that passed a severe test was corroborated. He formalized that statement with this equation that applies his concept of logical probability. The equation asks how much more logically probable the evidence is because of the hypothesis as opposed to background knowledge alone	
Bayes's formula [$p(h, e) = p(e, h) \times p(h)/p(e)$]	
$p(h, e)$	Posterior probablity of a hypothesis given some evidence
$p(e, h)$	Probability of the evidence given a hypothesis; proportional to likelihood
$p(h)$	Prior probability of a hypothesis
$p(e)$	Probability of the evidence
In Bayes's formula the probability terms on the right side of the equation are degrees of belief, rather than frequency probabilities, similar to Popper's logical probability. Comparing the likelihood function in this formula, $p(e, h)$, to the first term of Popper's corroboration formula, $p(e, hb)$, the only observed difference is the presence of b in the corroboration formula. Presumably, background knowledge would assert itself when determining the probabilities in Bayes's formula	

Second, they emphasized Popper's writings in which the philosopher suggested that statistical algorithms/measures—most importantly, Fisher's likelihood function—underlined his thinking on corroboration (and falsification). Specifically, after claiming that statistical methods are hypothetico-deductive in nature, Popper states "We can interpret ... our measure of degree of corroboration as *a generalization of Fisher's likelihood function*; a generalization which covers cases such as a comparatively large δ, in which Fisher's likelihood function would become clearly inadequate. For the likelihood of h in light of the statistical evidence e should certainly not reach a value close to its maximum merely because (or partly because) the available statistical evidence e was lacking in precision." (Popper, 2002, p. 432, his italics)

It is notable that Siddall and Kluge do not mention this passage; however, it is just as notable that De Quiroz and Poe do not mention Popper's earlier statement "Thus we have proved that the identification of degree of corroboration or confirmation with probability (and even with likelihood) is absurd on both formal and intuitive grounds ..." (Popper, 2002, p. 407). Given Popper's outlook as detailed in the main text of *The Logic of Scientific Discovery* written in the 1930s, before the appendixes from which the two quotes above were taken,[3] it is odd that he would rely so heavily on any probabilistic approach to the logic of science. Perhaps Popper found himself on a slope: needing a complement to falsification, he wrought corroboration; needing an explanation of corroboration, one distinct from 'truth confirmation,' he slid down into probabilistic reasoning, ending up with the corroboration formula discussed above (and see Table 1). In doing so, he left behind plenty of ammunition for both cladists and likelihoodists. (See Faith and Cranston, 1992 for a discussion of the use of Popper's corroboration and logical probability in cladistics without appeals to the formula.)

Having used some of this ammunition to defend ML, De Queiroz and Poe (2001) counter-attacked cladistic parsimony. "Unlike likelihood methods, parsimony methods are not based on explicit probabilistic models and thus they provide no basis for translating the minimum number of character transformations required by a tree into the probability of the observed distribution of character states among taxa given a tree." De Quiroz and Poe argue that using a methodology such as cladistics that includes only descent with modification as background knowledge (b) (e.g., Kluge, 1997), it is not possible to assign values to Popper's terms $p(e, hb)$ and $p(e, b)$. More generally, this observation illustrates how difficult it is to conform strictly to the minutiae of Popperian philosophy. An astute reader may note that the different viewpoints of cladists and likelihoodists discussed here mostly ignore earlier objections (see above) to applying Popperian philosophy to phylogenetic systematics.

The relative merits of cladistic parsimony and ML need to be examined in a new, Popper-free light. Here are a few questions that deserve further attention: one concern about likelihood is the use of frequency probability/statistics to infer unique historical events. Felsenstein (2003, pp. 144–145) has questioned this critique, stating "suppose we toss a coin 100 times and get 58 heads. We can regard the experiment as repeatable and infer the probability of heads. But suppose that,

[3] It seems pointed that Popper states in the note to these appendixes from the 1959 edition "... I [can] still agree with almost all the philosophical views expressed in [The Logic of Scientific Discovery (1935)], and even with most of those on probability—a field in which my ideas have changed more than any other ... ."

after we finish tossing, the coin rolls to the floor and then down a drain and disappears forever. Are not the 100 tosses now historical singularities? Yet clearly nothing important has changed that prevents us from inferring the probability of heads!" While the point regarding a statistical assessment of past events is taken, is the example of a coin flip (with one of two possible outcomes per flip) truly equivalent to reconstructing the phylogenetic relationships of N taxa (N usually $\gg$ than 2)? Are speciation events to be considered equivalent to coin flips? If so how many possible outcomes are there?

When comparing the relative merits of the two methodologies, explanatory power, in a general sense, is worth consideration. What inferences can we make once we have found an optimal tree? In the case of cladistic parsimony, both synapomorphies and homoplasies may be discovered in the data, allowing the formation of evolutionary hypotheses. In the case of ML, what new inferences about the data can be made, if any?

Despite unresolved philosophical questions regarding the merits of statistical phylogenetic inference, a new methodology has branched off from ML: Bayesian inference of phylogeny.

4. Popper Passé?

. . . I do not believe that my definition of degree of corroboration is a contribution to science except, perhaps, as an appraisal of statistical tests.

Popper (1983)

4.1. Bayesian analysis

A new era of phylogenetic statistical inference is now underway. Bayes's theorem is a long-standing formula for calculating conditional probabilities. Though a Bayesian approach to phylogenetics was first suggested over 30 years ago (Farris, 1973), it has only more recently come into practice. Instances of Bayesian inference of phylogenies (reviewed in Huelsenbeck et al., 2001 and Holder and Lewis, 2003) have been on the rise in the last 5–10 years, and show no signs of slowing down. Bayesian analysis is a probabilistic measure that combines prior probabilities of hypotheses and data with a likelihood measure yielding a posterior probability. It is formally written

$$p(h,e) = \frac{p(e,h) \times p(h)}{p(e)},$$

where the first term of the numerator is the likelihood function, the second the prior probability of the hypothesis, where the denominator is the probability of the evidence, and where the result is the posterior probability. Proponents of Bayesianism do not view probability as the possible outcomes of chance events, but rather as degrees of belief (Sober, 1988).

In Bayesian phylogenetics, the likelihood typically is calculated using Markov models of character evolution, and the prior probability of the phylogenetic hypotheses are equal (Huelsenbeck et al., 2001). It has been noted (Sober, 1988) that when the values for $p(h)$ and $p(e)$ are the same, different values of $p(h, e)$, the posterior probability, are derived only from differences in $p(e, h)$, the likelihood. Nonetheless, unlike ML, Bayesian analysis can utilize complex parameter rich models, yet yield fast results (Holder and Lewis, 2003).

4.2. Popper and probabilistic phylogenetic methods: Bayesian analysis

Was Popper a Bayesian? Given that the Bayesian approach is clearly inductive, it is at odds with Popper's philosophy as described in his primary work, *The Logic of Scientific Discovery* (Popper, 2002). Bayesian inference is an evidence-relationship, or confirmationist, approach, and Popper's corroboration is a non-Bayesian testing approach to the evaluation of hypotheses (Mayo, 1996). However, Popper's foray into probability in his search for a definition of corroboration sent him in a direction that can be construed as Bayesian/likelihoodist (De Queiroz and Poe, 2001, 2003; Felsenstein, 2003). This foray, embodied in Popper's corroboration formula discussed above undermines our ability to see Popper only as deductivist/falsificationist. It is interesting to note, though, that the formula is only discussed in the systematics literature reviewed herein, not the philosophical literature. Ironically, the search for real meaning in Popper's corroboration concept—a reason why passing a severe test makes a particular hypothesis special—has taken systematists into the most paradoxical and confused area of Popper's philosophy.

Consider the following statement (Sober, 1988, p. 180): ". . . there is no such thing as 'accepting' hypotheses at all. All that one does in science is assign degrees of belief . . . ." This sounds vaguely Popperian, but consider the continuation of this thought: "according to this position, the scientific evaluation of hypotheses proceeds in accordance with Bayes's theorem." The difference between Popper and Bayes is that the Popperian philosophy does not allow for assignment of 'degrees of belief.' This long-standing philosophical problem of the significance of non-falsification in Popperian testing applies to systematics: why should we accept any given tree as the best one? Whether or not having a high posterior probability is the answer to that question remains dependant on logical justification of the Bayesian inference approach to systematics.

4.3. The philosophical future of systematics

Rieppel (2003) recently questioned whether Popper's philosophy of science is relevant to systematic methodology and suggested that systematics should operate "under the influence of the philosophy of its own exponents." A philosophy of systematics founded by systematists need not completely ignore Popperian philosophy of physics. The Popperian spirit or critical attitude toward hypotheses is fundamental to all science. As Popper warned, if one does not look for faults in one's own hypothesis, someone else will. But in what ways can a philosophy of systematics be built independent of Popper? Or for that matter, can we construct a philosophy of systematics without appealing to any philosopher whose thinking is ignorant on the subject of systematics or biology in general? To date, philosophy has been the horse pulling the science cart in systematics' journey to find its logical justification. It might be timely to reverse roles and allow science to play the part of the horse pulling the cart of philosophy to finish the journey begun decades ago by the dedicated workers interested in making logical sense of their science.

Acknowledgments

The authors would like to the Roberto Keller, Kurt Pickett, and Ilya Temkin for comments and criticisms on an earlier draft of this paper.

References

Bartley III, W.W. 1983. Editor's foreword. In: Realism and the Aim of Science. Rowman and Littlefield, Totowa, New Jersey, pp. xi–xv.

Bock, W.J., 1973. Philosophical foundations of classical evolutionary classification. Syst. Zool. 22, 375–392.

Brower, A.V.Z., 2000. Evolution is not a necessary assumption of cladistics. Cladistics 16, 143–154.

Cracraft, J., 1978. Science, philosophy, and systematics. Syst. Zool. 27, 213–216.

De Queiroz, K., Poe, S., 2001. Philosophy and phylogenetic inference: a comparison of likelihood and parsimony methods in context of Karl Popper's writings on corroboration. Syst. Biol. 50, 305–321.

De Queiroz, K., Poe, S., 2003. Failed refutations: further comments on parsimony and likelihood methods and their relationship to Popper's degree of corroboration. Syst. Biol. 52, 352–367.

Edwards, A.W.F., 1992. Likelihood. The Johns Hopkins University Press, Baltimore.

Faith, D.P., Cranston, P.S., 1992. Probability, parsimony, and Popper. Syst. Biol. 41, 252–257.

Farris, J.S., 1973. A probability model for inferring evolutionary trees. Syst. Zool. 22, 250–256.

Franklin, J., 2001. Resurrecting logical probability. Erkenntnis 55, 277–305.

Felsenstein, J., 2003. Inferring Phylogenies. Sinauer Associates, Sunderland, Massachusetts.

Gaffney, E.S., 1975. A phylogeny and classification of the higher categories of turtles. Bull. Amer. Mus. Nat. Hist. 155, 391–436.

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: Traditional Bayesian approaches. Nat. Rev. Genet. 4, 275–284.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Hull, D.L., 1980. Karl Popper and Plato's metaphor. In: Platnick, N., Funk, V.A. (Eds.), Advances in Cladistics, vol. 1. Columbia University Press, New York, pp. 177–189.

Kitts, D.B., 1977. Karl Popper, verifiability, and systematic zoology. Syst. Zool. 26, 185–194.

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

Kluge, A.G., 2001. Philosophical conjectures and their refutation. Syst. Biol. 50, 322–330.

Lakatos, I., 1978. The Methodology of Scientific Research Programs. In: Worrall, J., Currie, G. (Eds.), Cambridge University Press, New York.

Mayo, D.G., 1996. Error and the Growth of Experimental Knowledge. The University of Chicago Press, Chicago.

Mishler, B.D., 1999. Getting rid of species?. In: Wilson, R.A. (Ed.), Species New Interdisciplinary Essays. The MIT Press, Cambridge, Massachusetts, pp. 307–315.

Nelson, G., 1978. Classification and prediction: A reply to Kitts. Syst. Zool. 27, 216–218.

Patterson, C., 1978. Verifiability in systematics. Syst. Zool. 27, 218–222.

Platnick, N.I., Gaffney, E.S., 1977. Book review. Syst. Zool. 26, 361–363.

Popper, K.R., 1964. The Poverty of Historicism. Harper Torchbooks, New York.

Popper, K.R., 1965. Conjectures and Refutations: The Growth of Scientific Knowledge. Harper Torchbooks, New York.

Popper, K.R., 1974a. Darwinism as metaphysics. In: Schlipp, P.A. (Ed.), The Philosophy of Karl Popper, Book I. Open Court, La Salle, Illinois, pp. 133–143.

Popper, K.R., 1974b. Replies to my critics. In: Schlipp, P.A. (Ed.), The Philosophy of Karl Popper, Book II. Open Court, La Salle, Illinois, pp. 961–1197.

Popper, K.R., 1979. Objective Knowledge: An Evolutionary Approach. Oxford University Press, New York.

Popper, K.R., 1980. Letter to New Scientist 21 August.

Popper, K.R., 1983. Realism and the Aim of Science. Rowman and Littlefield, Totowa, New Jersey.

Popper, K.R., 2002. The Logic of Scientific Discovery. Routledge Classics, New York.

Putnam, H., 1974. The "corroboration" of theories. In: Schlipp, P.A. (Ed.), The Philosophy of Karl Popper, Book I. Open Court, La Salle, Illinois, pp. 221–240.

Rieppel, O., 2003. Popper and Systematics. Sys. Biol. 52, 271–280.

Ruse, M., 1977. Karl Popper's philosophy of biology. Philos. Sci. 44, 638–661.

Ruse, M., 1979. Falsifiability, consilience, and systematics. Syst. Zool. 29, 530–536.

Salmon, W., 1967. The Foundations of Scientific Inference. University of Pittsburgh Press, Pittsburgh, Pennsylvania.

Settle, T., 1979. Popper on "When is a science not a science?". Syst. Zool. 28, 521–529.

Siddall, M.E., Kluge, A.G., 1997. Probabilism and phylogenetic inference. Cladistics 13, 313–336.

Sober, E., 1988. Reconstructing the Past. MIT Press, Cambridge, Massachusetts.

Stamos, D.N., 1996. Popper, falsifiability, and evolutionary biology. Biol. Philos. 11, 161–191.

Wiley, E.O., 1975. Karl R. Popper, systematics and classification: A reply to Walter Bock and other evolutionary taxonomists. Syst. Zool. 24, 233–243.

Glossary

Background knowledge: any knowledge accepted, even if only tentatively, as unproblematic in a test of a hypothesis.

Bayes theorem, Bayesianism: a theorem describing how the conditional probability of a set of possible causes for a given observed event can be computed from knowledge of the probability of each cause and the conditional probability of the outcome of each cause.

Consilience: concurrence of inductive inferences.

Confirmation: additional proof that something that was believed (some fact or hypothesis or theory) is correct.

Conjecture: reasoning that involves the formation of conclusions from incomplete evidence.

Corroboration: passing a severe test.

Critical test: an evaluation that can easily be failed.

Deduction (Deductivism, Deductivist): reasoning in which premises follow from a general statement; doctrine advocating use of deductive reasoning; a follower of deductivism.

Degree of confirmation: a measure of the strength of additional proof that something that was believed (some fact or hypothesis or theory) is correct.

Degree of corroboration: the degree to which a hypothesis has stood up to severe tests.

Empirical content: the sum or range of what has been perceived, discovered, or learned derived from experiment and observation rather than theory.

Empiricism: doctrine of pursuit of knowledge by observation and experiment.

Epistemology: branch of philosophy dealing with knowledge, how it is gained, and its relationship to reality.

Evidence: anything presented to the senses and offered to demonstrate the existence or non-existence of a phenomenon.

Explanation: that which makes known the logical development or relationships of phenomena.

Falsification (Falsificationism, Falsificationist, Falsifier): the act of disproving a theory; the school of thought that emphasizes the importance of falsifiability as a philosophical principle; one who proceeds in explaining the world by disproving theories; piece of evidence that disproves a theory.

Fisher's likelihood function: a statistical statement based on the idea that the "best" parameter value will be the one which maximizes the likelihood that the experiment would have turned out the way it actually did; any parameter values can be chosen, but some will be more likely to be the true values, given the experimental results.

Frequency probability: numerical probability based on mathematical observation. The 1/6 probability that a '5' will result from the roll of a single die is an example of frequency probability.

Historical inference: a deduction or induction from premises about history.

Historicism (sensu Popper): the view that history is governed by necessary laws of development.

Hypothesis: a tentative theory about the natural world, potentially able to explain certain facts or phenomena.

Improbability of evidence: a measure of the unlikelihood of something presented to the senses and offered to demonstrate the existence or non-existence of a phenomenon.

Induction (Inductivism, Inductivist): reasoning from specific instances to a general conclusion; doctrine advocating use of inductive reasoning; a follower of inductivism.

Likelihood (Likelihoodist): mathematical concept in which the likelihood, L, of a hypothesis, h, given data, e, and a model, b (defined below), is proportional to the probability of e given h times an arbitrary constant. $L(h, e) \approx p(e, hb)k$; an advocate of using likelihood to test hypotheses.

Logic: the set of principles governing reasoning.

Logical probability (sensu Popper): the converse of testability.

Metaphysics: empirically untestable ideas about the physical world—a set of beliefs about the nature of reality. Metaphysical ideas may become scientific once they are testable, for example, atomic theory.

Model: a mathematical description of an empirical system.

Numerically universal statement: a conjunction of singular statements, i.e., a finite set, that is thus enumerable.

Operationalism: the doctrine that value of a hypothesis consists of the operations involved in testing or applying it.

Parsimony: (1) principle of preference of the most economical explanation. Sometimes referred to as Occam's razor after William of Occam (1285–1349) who championed the principle in his writings against the papacy (2) criterion for choosing among competing phylogenetic hypotheses in which the preferred tree is the one with the fewest character state changes.

Probability (Probabilism): generally, the chance that an event will occur. Various approaches to considering or calculating probability exist, e.g., frequency probability; doctrine advocating the use of probability to solve (generally) scientific questions. May apply to any form of probability including frequency or logical probability.

Refutation: the demonstration of falsehood via evidence of test.

Relevance: pertinence to a given issue or question.

Retrodiction: inferential reasoning about the past from data from the present.

Singular or individual statement: a highly restricted statement with respect to time and space, a finite set.

Testability: the characteristic of being able to be corroborated; the more testable a hypothesis is the better it can be corroborated.

Universal statement: statement pertaining to that which exists or functions in all places and all times without exception.

Verification (Verificationism, Verficationist): confirmation establishing truth; doctrine of conducting inquiry by seeking to verify one's ideas or hypotheses; an advocate of verificationism.