Biology and Philosophy (2005) 20:489–515
DOI 10.1007/s10539-004-0757-2

Outline of an explanatory account of cladistic practice

NICO M. FRANZ
Department of Entomology, Comstock Hall, Cornell University, Ithaca, NY 14853, USA; (E-mail: nmf2@cornell.edu; phone: +607-255-8050; fax: +607-255-0939)

Received 2 October 2003; accepted in revised form 17 June 2004

Key words: Classification, Congruence test, Evolutionary homeostasis, Homology, Natural kinds, Parsimony, Phylogeny, Projectibility, Reference, Taxonomy

Abstract. A naturalistic account of the strengths and limitations of cladistic practice is offered. The success of cladistics is claimed to be largely rooted in the parsimony-implementing congruence test. Cladists may use the congruence test to iteratively refine assessments of homology, and thereby increase the odds of reliable phylogenetic inference under parsimony. This explanation challenges alternative views which tend to ignore the effects of parsimony on the process of character individuation in systematics. In a related theme, the concept of homeostatic property cluster natural kinds is used to explain why cladistics is well suited to provide a traditional, verbal reference system for the evolutionary properties of species and clades. The advantages of more explicitly probabilistic approaches to phylogenetic inference appear to manifest themselves in situations where evolutionary homeostasis has for the most part broken down, and predictive classifications are no longer possible.

Why use parsimony in systematics? In this paper I present certain elements of cladistic practice in a naturalistic philosophical context. There are two major and interrelated themes. On the one hand, my intent is to *explain* how a particular implementation of the parsimony criterion – the so-called congruence test – is used iteratively to achieve (more) phylogenetically informative observations of similarity. To this end, I interpret statements of homology as natural kind terms in the Quinean sense. The congruence test is portrayed as a means to inductively assess and increase their projectibility. I contend that parsimony considerations affect the process of character selection in individual analyses, and also contribute to the long-term entrenchment of a phylogenetic terminology. This position challenges other accounts that are primarily concerned with the performance of parsimony algorithms *vis-à-vis* a range of evolutionary process models (Sober 1988). A more complete picture would emerge if the observation-refining effects and process-related assumptions of cladistic parsimony were viewed as complementary parts of an increasingly powerful inferential package. In short, cladists may exploit the congruence test to select character systems and interpretations that increase the odds of reliable phylogenetic inference.

On the other hand, I use the concept of homeostatic property cluster natural kinds (Boyd 1999) to reassess the strengths and limitations of the cladistic approach, in relation to alternative inference methods. My preferred measuring stick for this evaluation is the age-old task of systematics to provide a verbal reference system for the natural, evolutionary properties of species and clades. I suggest that slow to moderate rates of evolution are necessary for such reference achievements to take place. The phylogenetic inference capabilities of parsimony likely exceed these requirements. This means that cladistic practice is adequate in principle for the task just mentioned. Whenever the homeostasis of evolutionary property clusters has broken down, the inference of phylogenetic relationships may still be possible using (also) more explicitly probabilistic methods. Yet from a traditionalist perspective, this comes at the cost of individuating clades only in reference to non-verbal, largely ambiguous statements of phylogenetic similarity. There is an evolutionary threshold beyond which a cladistic notion of homology seems no longer appropriate. The concept of homeostasis is not only useful for characterizing this threshold – naturalistically – but also for understanding a common trend in molecular phylogenetics away from predictive, property-centered classifications.

Preparing the stage

For the most part this paper has been written for philosophers of science. A fair amount of space is used to explain various conceptual and technical aspects of cladistic practice, and to introduce the reader to the relevant topics of debate. More prior knowledge is assumed about concepts like natural kinds and projectibility. Nevertheless, from time to time I will address how the present account may be reconciled with positions commonly advocated by cladists. These paragraphs are marked accordingly.

The current section has the following main functions. It situates the naturalistic, explanatory approach towards systematics as one among many worthwhile philosophical projects. It also reiterates an important assumption – possibly a lesson – about the only approximate relation of certain inferential practices in science and scientists' conceptual accounts of their underlying structures. After that the reader is introduced to the sequence of points and arguments which will follow.

Explaining the achievements of cladistics, in terms of its contribution to reliable propositions of phylogenetic homology and relationship, is one of the several approaches that philosophers of science can use to advance their issues. Alternative projects may evaluate the structure of this research tradition in comparison to other historical and experimental sciences, or with respect to attractive philosophical theories. Which theory can claim more support in light of what cladists seem to (or should) do? Understanding the assumptions inherent in particular inference schemes, such as parsimony, is a related task

(Sober 1988). If new insights are presented about competing methods, this can have implications for practice. Philosophers are also often interested in the historical and sociological conditions that produce and maintain research traditions like cladistics (Hull 1988). Depending on their personal backgrounds and motivations, explanations of the inferential successes of cladistic practice may or may not play a central role. They do in present account.

For a variety of reasons, cladists themselves are probably even less concerned with explanations of inferential achievements. This should be expected in particular if those explanations involve inductive and probabilistic arguments about the reliability (or even approximate truth) of particular cladistic propositions, which may seemingly undermine a critical attitude towards induction and truth-like statements. That attitude, in turn, has often served to portray cladistics as a respectable discipline, and thereby distinguish it from other schools within systematics. Philosophically, cladists have been more interested in themes like demarcation and justification than in explaining success.

The former paragraph gradually moves towards my second cautionary point about the inexact relation of scientific practices and practitioners' conceptual accounts of them. As mentioned above, scientists may adopt a philosophical stance partly because it preserves the coherence of an incipient research community (Hull 1988). Positions that turn out indefensible in the ultimate philosophical analysis can nevertheless persist and thrive in science for considerable periods of time. If and to the extent that these incorrect conceptions are vague enough, or loosely implemented, they can even be guides to good, reliable practice. This is a well-known phenomenon, and a persistent challenge for philosophers of science who must sometimes ignore what is said, while taking a closer look at what is done. Case in point: for many years pre-Darwinian taxonomists were regarded as more or less rigid executioners of Aristotelian essentialism, understood in a narrow sense that conflicts with evolutionary thinking. As a perceptive analysis of their actual classificatory methods and schemes reveals, however, they adopted these concepts pragmatically, or even abandoned them to increase the naturalness of taxa (Winsor 2003). They were non-essentialists in practice.

Philosophers of science who analyze the ins and outs of cladistic practice have similar challenges ahead of them. It is probably less enlightening to read much of the cladistic literature than to sit next to a cladist when she runs the most recent version of her character matrix on a parsimony-implementing software application, and obtains a completely unresolved consensus of many cladograms, or one that apparently makes 'no sense.' If it were only for the resulting publication, one may never know what kinds of inferences and assumptions were made at that moment. Worse, the results could be cast in an inadequate philosophical context that further clouds the picture. A lot of effort can be spent just to sort out these misrepresentations, without seeing much of the underlying approach.

My explanatory account is particularly concerned with those aspects of cladistic practice that – for one reason or another – are less apparent from the

outside. Perhaps there is a lesson here, more general than anything which follows. Considering the contentious history of cladistics, it seems at least possible that philosophers of science will have to understand certain inference patterns *in spite of* what cladists say.

Overview of argumentative structure

I will start this account by outlining what kinds of cladistic achievements are in need of explanation. What should count as inferential success in the present context? Then a particular version of the overall aims of cladistics is presented. The philosophical conceptions of natural kinds and homeostatic property clusters are introduced in close relation to these aims. They will be relevant to each of the two major themes.

The initial sections of the first theme turn more technical. I ask what sorts of conditions must be met for a matrix of characters (pertaining to a set of taxa) to confer information about phylogenetic similarity and relationship. This leads up to the central argument; that the parsimony-implementing congruence test is used iteratively (and inductively) to assess and increase the compliance of a character matrix with the observation- and process-related conditions necessary for reliable inference of phylogeny. Over longer time periods, the congruence test also contributes to the entrenchment of a reliable phylogenetic terminology. This position is viewed as a challenge to alternative projects which place less emphasis on the effects of parsimony on the processes of character individuation and refinement.

The second, shorter theme examines the nature of cladistic inferential and referential practice in light of the concept of homeostatic property clusters. There is a controversy within contemporary systematics over the strengths and limitations of cladistics, in comparison to more explicitly probabilistic methods. These issues come up at various stages throughout the paper, and are addressed more openly towards the end. The concept of evolutionary homeostasis offers a new way for characterizing the competing approaches, and illustrates why cladistics is particularly well suited for traditional, property-based descriptions of clades.

What needs explanation?

It is appropriate to first describe what sorts of inferential achievements related to cladistic practice are worth explaining. Is it even possible to tell successful phylogenetic hypotheses from unsuccessful ones? Clearly one can review the recent history of systematics – after the Hennigian approach had been combined with powerful software applications – and witness a change of pace. Cladograms have not only increased in number and taxonomic range, but also (for the most part) in evidential support and precision. There is accelerating progress towards proposing a phylogeny of life. Meanwhile, recurrent conflicts

among alternative hypotheses of relationship are informed, complicated, or resolved by additional relevant evidence. It could seem as though research is thriving. Yet at the same time, systematists do not work experimentally (Cleland 2002), and the phenomena they investigate are not as fundamental as the laws of physics (Rosenberg 1994). Whether science has been successful is immediately obvious when it comes to constructing an airplane, but perhaps less so in the case of classifying millions of insects.

Like all valid scientific propositions, cladistic hypotheses have to be *projectible* (Goodman 1954). The separation of projectible from unprojectible statements of homology and phylogeny can only succeed, however, if there is a sufficiently discriminatory body of projectible overtheories. The problem is that many of the currently available ideas about regularities and constraints in evolutionary processes do not facilitate such discrimination among alternative positions. Initially a particular series of character transformations in insects might be as plausible as another, considering what (little) is known about insect evolution. The absence of strong projectibility tests explains in part why cladists place much emphasis on observational evidence, and criticize the '*ad hoc*' use of presumably non-projectible 'stories' about the evolutionary (and thus phylogenetic) significance of adaptive or otherwise relevant characters. Such practice is considered to be excessively speculative.

The impossibility – if only temporary – of assessing the projectibility of alternative cladistic hypotheses sometimes makes it more reasonable to view cladograms as mere summaries of observations. This deflationary attitude is a reflection of real epistemic limitations of cladistic practice, and also of the present explanatory account. Projectibility is after all a successively emerging property of theories. In many cases it is too early to make a judgment. Still, one has to acknowledge the very significant contributions that cladograms make to hypothesis testing in other biological disciplines, especially those involving questions about character phylogeny. Examination of a recent journal of evolution, in comparison to one that is three decades old, will confirm the impact of this 'phylogenetic revolution.' The advancement of knowledge in related areas of investigation represents an achievement of cladistics that calls for an explanation.

Another, not unrelated mark of success is the entrenchment of particular cladistic hypotheses in the course of time, as additional evidence is accumulated in support of them. This kind of reliability, or inductive efficiency, would be unexpected unless the original analyses identified characters that are in congruence with other, unobserved characters. The epistemic causes for congruence, and the emergence of a precise and sufficiently stable taxonomy for the various users of cladistic research, need to be explained.

Natural kinds and the aims of cladistics

Cladistic practice has been associated with a range of philosophical conceptions and motivations (Ruse 1979). In this paper I adopt a naturalistic,

scientific realist interpretation. According to this view, cladists in the Hennigian tradition (Hennig 1966; Farris 1983; Rieppel 1988) are concerned with elaborating hierarchical summaries and classifications for the historical products of evolution. Cladograms are rooted in extensive analyses of observable characters. Typically cladists take themselves to be successors of the Linnean tradition. The aim is to provide a taxonomy – constituted by names and diagnoses for species and clades – for the organismal world. I emphasize these language-related aspects early on because of their relevance to the second theme. There are other research traditions where inferences of phylogenies are also central, though not always translated into classifications.

In a very inclusive sense, cladists work on a reference system to facilitate a fruitful communication about the properties of taxa among the biological disciplines. In addition to identifying individual species, they propose homology statements among sets of species. The latter are tested for their mutual congruence (roughly, whether their presence and absence corresponds to the overall hierarchical structure) with parsimony; and then optimized (mapped) along the internal branches of the phylogenetic tree. Again, the intention is to produce an efficient observational terminology that reflects an apparent hierarchical order in nature. The prospect of combining diagnostics and evolutionary characterizations is perceived as a strength (Hennig 1966).[1]

From a realist perspective, the history of cladistics may be interpreted as a gradual alignment process between a scientific terminology on one side, and hierarchically structured phenomena in nature on the other side. Conflicts and solutions emerging within this paradigm are comprehensible as attempts to establish and improve a reference relation between homology statements and the corresponding evolutionary regularities that have shaped the living world.

If homology is regarded as an evolutionary phenomenon, then the historical unfolding of phylogenetic research can be elucidated with a naturalistic account of natural kinds (Quine 1969).[2] The earliest attempts to produce a natural system used a largely unspecified notion of similarity among organisms. It included analogy and ecological similarity. Throughout the centuries, this notion has been refined and transformed into an increasingly theoretical one. Some of the critical refinements are reflected in the separation of homology from analogy and homoplasy (convergence, reversal); of ontogenetic

---

[1]To clarify an underlying assumption: cladists can argue at length whether 'evolutionary theories' are, or should be, involved in the process of character analysis. Yet cladistic hypotheses are used by evolutionarily oriented researchers regardless of where one is situated along the spectrum of positions. Moreover, it is not unusual to observe 'agnostic cladists' opposing certain (excessively adaptational) interpretations of characters in ways that are actually evolutionarily plausible. Therefore my explanatory approach will center around the contributions which cladograms and character phylogenies make to an understanding of evolution.

[2]Quine (1969, p. 136) mentions how a sophisticated, phylogenetic concept of similarity would confer fishes natural kind status 'in the corrected, whale-free sense.' This argument has been further developed by Griffiths (1994) who provides various insights regarding the naturalness of clades and homologies that are adopted here.

(only developmental) from phylogenetic (also historical) identity; in the exclusion of paralogy ('lateral' replication within genomes) and xenology ('lateral' transfer among genomes; Mindell and Meyer 2001); and in the inclusion of relevant, vertically transmitted behavioral, biochemical, and epigenetic similarities. None of these refinements of phylogenetic homology are very intuitive.

It is also useful to think of the aforementioned transformations in the Quinean sense with an eye for future developments. Presumably at some point systematists will refer to the diagnostic attributes of certain clades not only as synapomorphies mapped along cladograms, but also as causally sustained phenomena at the molecular level (Brigandt 2002). Perhaps at this juncture they will have achieved as much referential precision as possible in light of the complexity and contingency of evolution.

Recently Boyd (1991, 1997, 1999) has further developed and expanded the theory of natural kinds. The notion of homeostatic property clusters is instrumental in explaining reference to evolving, non-eternal, spatiotemporally contingent instances of kinds. It posits the – typically imperfect – co-occurrence of properties which have a tendency to be clustered together in nature because of the presence of underlying causal mechanisms. Terms that refer to these clusters have 'only' *a posteriori* definitions, which will have to specify the relative significance of the various properties. As a consequence of the imperfect causal association, the terms will necessarily exhibit some extensional interdeterminacy. Homeostatic property cluster kinds are individuated like historical objects. Their naturalness is reflected in the contribution that reference to them makes to the accommodation of successful inferences in the respective scientific traditions. Boyd (1999) regards evolving species, possibly also clades, as paradigmatic examples of such kinds. The second theme adopts this interpretation, and explores how varying levels of evolutionary homeostasis limit our ability to individuate taxa using cladistic or more explicitly probabilistic approaches.

Requirements for a phylogenetically informative character matrix

Modern cladistic practice revolves around the task of making relevant observations on a set of taxa, and translating these into a matrix of characters and character states ('0s' and '1s'). Even though the analysis of such a matrix (to obtain a cladogram) is not trivial, much of what cladists achieve is manifested in the earlier, observation-encoding stages. I will now review what sorts of conditions must be met to render a matrix phylogenetically informative.

Observation in science is widely acknowledged to be theory-laden. Starting at least with Aristotle (Lennox 1991), the history of systematics reaffirms this conclusion. Significant theoretical refinements of what counts as a valid observation in taxonomy include the works of Belon, and especially of

Geoffroy Saint-Hilaire whose *principe des connexions* embodies the still relevant criterion of topographical correspondence (whether structures resemble each other in position and arrangement; Rieppel 1988). Hennig (1966) argues explicitly that systematists should not aspire to extract all imaginable kinds of properties from taxa, but only those conferring phylogenetic information. Separating signal from noise is critical to phylogenetic inference. In actual practice, this translates into the challenge of observing characters that might contain conflicting evidence (yet apply to the problem complex) from those that are simply not informative at the intended level of research. Cladists will typically refer to this as the requirement to observe and report all critical evidence (Kluge 1989).

One can review the theory-laden, selective aspects of observation in cladistics from a more technical perspective, by listing the requirements for obtaining a character matrix that encodes a reliable phylogenetic signal.[3] Many of these criteria will be revisited at a later stage.

Discreteness – this is a theoretical as well as a practical (algorithmic) condition. Tokogenetic variation among populations, continuous characters, and certain relations from morphometric or genetic distance analyses are likely to be removed from the matrix. *Heritability*; the included properties should be replicated and transferred vertically in the course of evolution through the process of descent with modification. Exclusively ontogenetic or extrinsic attributes are omitted. *Heterogeneity* – deceptive biases in the abundance and directionality (distribution throughout the tree of life) of homoplasy should be eliminated. *Hierarchy*; the characters have to indicate orthologous (vertical) phylogenetic relationships. This calls for the exclusion of information related to phenomena like paralogy, xenology, and reticulation. *Independence* – each observation should correspond to an evolutionarily 'separate' part of evidence. These parts can then be subjected to reciprocal tests as independent indicators of phylogenetic relationship. *Precision* – the translation (process of encoding) of the original homology assessment to the numerical circumscription must be accurate. This require cladists to make judgments as to whether a particular sequence of character state transformations is more plausible than others (additive versus non-additive multistate characters), or whether certain properties are 'inapplicable' (what homologous shape have the wings in secondarily wingless insects?). The alignment of DNA sequence data presents a similar problem. *Transformation* – the extent of variation among character states has to be adequate for the level of analysis. Invariable character systems, or ones that evolve at rates much higher than the rate of cladogenesis, are excluded. Neither can confer an informative signal for the time period during which the examined taxa evolved. *Weight* – the weight assigned to a character should approximately

[3]This list is neither exhaustive nor without redundancy. The significance of each criterion can vary from one character system to another. Some are (expectedly) controversial among cladists. They are intended to capture an overall tendency.

correspond to the 'quantity' of phylogenetic evidence it represents. This can mean assigning proper weights to serially homologous attributes. For instance, a transformation on the legs of a millipede would not be upweighted by the number of structures on which it can be observed.

Although there will always be complications, a matrix that complies with the previous requirements is considered more likely to carry a reliable phylogenetic signal. The next step in my analysis involves asking to what extent these conditions are addressed by the *operational criteria* that systematists use to make relevant observations. Then I review the assumptions entailed in using *parsimony* as an algorithm to produce a hierarchical structure among them.

From observational criteria to causal mechanisms

The history of numerical taxonomy offers some idea as to what would occur if similarity were conceived as 'anything similar to an unbiased observer.' It is possible to extract an almost infinite number of traits from a set of organisms (such as 100 arbitrary variables characterizing the shapes of their leaves), and then classify them in a way that has no resemblance to phylogeny.

In cladistics, homologies among taxa are proposed in accordance with traditional criteria; topographical correspondence, special quality (such as an elaborate courtship display), and connection by intermediates (often apparent during the development of a specimen). However, the operational adherence to these criteria cannot actually *explain* why cladograms have phylogenetic relevance (Brigandt 2002). The algorithmic structuring of observations with parsimony does not automatically transform them into causally sustained natural kinds. To arrive at a convincing explanation, philosophers and developmental geneticists must instead provide a complex account that specifies how actual implementations of the traditional homology criteria succeed, and fail, at individuating properties of taxa that have such underlying mechanisms (Wagner 2001). A very coarse picture of how this could work is described below.

For obvious reasons, the activity of extracting characters from preserved specimens cannot immediately address transformational or stabilizing evolutionary processes. No systematist witnessed what (has) occurred when (and since) insects evolved wings. These epistemic limitations not withstanding, comparative studies have a rather impressive record of identifying properties with sufficient levels of homeostasis and uniqueness to confer reliable inferences of phylogeny. For this purpose, criteria like topographical correspondence appear superior to many plausible alternatives. The characterizations of 38,000 kinds of spiders by their spinnerets (silk-producing glands), or of 250,000 kinds of flowering plants by their endosperm (resulting from the process of double fertilization), are appropriate examples. These attributes have the 'smoking gun qualities' that Cleland (2002) portrays as a mark of

conclusive historical research. However, is it possible to *infer* – based on a comparative analysis of their presence and absence throughout the living world – that they are unique, homeostatic products of the same underlying causal mechanisms? Such an abductive inference is certainly not implausible. Testing it would involve the integration of various insights from evolutionary biology. For a start: it seems reasonable assume, in light of the phenotypic complexity of spinnerets, that their ontogenies and specific structures are controlled by a series of interacting genes. The observed morphological variation among all spiders suggests that many of these spinneret-regulating genes have undergone considerable evolutionary transformation. At the same time, particular DNA regions (responsible for coding the position and arrangement of spinnerets) must have remained relatively stable. When developmental geneticists identify these regions and investigate their role in the ontogeny and reproduction of spiders, it might well turn out that the phenomenon of spinnerets actually corresponds to a series of unique homeostatic mechanisms at the molecular level. The likelihood of this outcome indicates how powerful of the traditional homology criteria are for achieving such accommodation. Alternative criteria that may emphasize physical, chemical, ecological, or overall similarities are less likely to characterize clades and their causally sustained properties with this much accuracy.

Even without addressing the implementations of cladistic homology criteria, any account of how precise and reliable observations are obtained is already incredibly complicated (Rosch and Lloyd 1978; Maund 2003). Philosophers of science have yet to analyze the former activities in detail.[4] This is another important shortcoming in the present context. Clearly, if the original assessments of homologies among taxa are unable to provide at least an approximate signal of phylogeny, then there will be nothing to explain with respect to their structuring with parsimony. Cladistics is an empirical science, and so much of what it achieves must be related to the practical ability of cladists to make relevant observations.

Short of a deeper analysis of observation and representation in systematics, one can minimally state that homology propositions are in essence comparative and probabilistic. Established practice involves more sophistication than a simple-minded conjoining of the various traditional criteria for phylogenetic similarity. Homology, and the lack thereof, is not assessed with respect to an 'objective value,' but only in relation to the actually observed topographical and structural variations in organisms of this world ('genospace' and 'phenospace'). Whether a particular instance of transformation is judged significant, or not, depends on extensive studies of the variability of attributes within and among related species. Assessments of similarity are contextual. Their precision will increase with the exploration of the overall space of variation. This may explain why conflicting interpretations of homology among experienced systematists are often reflective of deeper

[4]Brady (1994) offers a curious phenomenological account of observation in systematics.

arguments. Far beyond disagreements about the (allegedly) same visual input, these conflicts evolve as a result of alternative and non-arbitrary evaluations of the relevant empirical and theoretical *context* for comparison.

The probabilistic nature of homology propositions is also expressed in a comparative setting. Cladists can sometimes study thousands of characters, pertaining to hundreds of species. To observe the exact congruence of several topographically unique and structurally complex attributes in a small subset of species would be an improbable event, *unless* they have a common origin. A classic example of the cognitive impact that such a signal can have on cladists is Platnick (1982), who argues that the presence of spinnerets and modified, sperm-transferring male pedipalps (appendages) in all spiders – and nowhere else – would indicate their naturalness even if no theory existed to explain such a coincidence of properties.

Identifying improbable associations of similarities in the overall space of variation, and then comparing likely explanations for their congruence, are paradigm activities of historical research. Cladists fit the description well enough. As suggested by Cleland (2002; though in another context), they will exploit the overdetermination of causes by their effects in these ways to assess the probability of homologous properties being manifested in different taxa.

Evolutionary process assumptions related to the use of parsimony

I am moving away from discussing systematists' abilities to make relevant observations to more general suppositions. Cladists analyze their elaborate character matrices with parsimony-implementing software applications (identifying the shortest trees). Nested sets of clades and character state transformations (from '0' to '1') are ultimately obtained by 'rooting' the results in the taxa that originated from the earliest events of cladogenesis ('outgroups'). At this point, it necessary to ask what the use of *parsimony* as an inferential standard assumes about the nature of evolutionary processes. As mentioned above there are other standards with diverging assumptions. In the past few decades, philosophers and systematists have gained substantial insights into these issues.

One core premise that systematists have to accept about the process of evolution is actually independent of any methodological preference. Sober (1988) makes a basic distinction between historical processes that preserve information, versus those which destroy it. In the latter case, for example if several unrelated evolutionary lineages converged from time to time on identical adaptive optima, inference of phylogeny will be impossible – regardless of the algorithm used. There is simply no valid signal preserved through space and time. The decision for or against a particular inference standard must be informed by its performance in cases where at least some information from the past has made it to the present.

Theoretical evolutionary process models and computer simulations have led to more precise evaluations of alternative methods.[5] Felsenstein (1978) conceived a scenario with a particular combination of unequal evolutionary rates in closely related lineages. Under such conditions parsimony would compute an incorrect topology – a phenomenon called long-branch attraction ('convergence outweighs common ancestry'). The reliable use of parsimony as a phylogenetic inference standard thus entails the assumption that this phenomenon was not predominant throughout the real history of life. Sober (1988) considered a more sophisticated model from which the following theorem was derived: synapomorphies (shared derived attributes) are more significant indicators of phylogeny than symplesiomorphies (ancestral similarities) if the former occur more frequently than the latter. Under the specific model conditions the inferential success of parsimony is independent of the abundance of homoplasy. This is what cladists had argued previously (Farris 1983), though not conclusively according to Sober (1988).

Recently, various process assumptions related to the use of parsimony have been translated into the language of the competing maximum likelihood inference approach. As summarized by Steel and Penny (2000), the topologies computed by either method will be identical under a particular set of conditions. These include: a Poisson model with equal probabilities for each possible character state transformation (from A to C, C to G, etc.), and 'no common mechanism' (potential complete independence) among evolutionary substitution rates within individual lineages. It is also possible to specify conditions that would render each method less accurate than the other, or even statistically inconsistent (worse performance as the amount of evidence increases). In the case of parsimony, these undesirable properties are minimized as the evolutionary rates decrease and invariable characters are added, or as the numbers of possible character states increase (Steel and Penny 2000). The latter conditions are probably the rule in morphological character matrices.

It is important in the context of evaluating the performance of alternative methods for phylogenetic inference to differentiate between necessary and sufficient assumptions for success (Sober 1988). Although parsimony has the tendency to compute correct topologies when there are continuously varying evolutionary rates or many character states, none of these conditions are required in all instances. Sober (2003) introduces the Akaike information criterion to make these relations more explicit. In short, one can investigate the accuracy of methods by examining whether their results have *ordinal equivalence* (provide a consistently structured answer; in this case an identical tree) under a range of evolutionary assumptions. Parsimony and maximum likelihood inference have ordinal equivalence under the model of 'no common mechanism.' Additional similarities and differences may be exposed in the future.

[5]This exploratory approach is analogous to philosophical inquiries about the truth conditions of statements in counterfactual situations.

What are the current lessons from the modeling approach? Certainly some tendencies are apparent. Character systems with excessive rates of evolution, combined with deceptive biases in the abundance of homoplasy and a paucity of states – are unlikely to yield accurate phylogenies when parsimony is used. The significance of these limitations should be proportional to the number of real life cases in which parsimony and maximum likelihood inferences are in conflict. In contrast, character systems evolving at continuously varied or slow (to moderate) rates, while transforming into many states, are often reliable indicators of phylogeny under parsimony. The effects of homoplasy on these relationships are not always easy to predict. Although parsimony will significantly underestimate evolutionary rates in many situations (analyses of DNA sequence data in particular), systematists are still confronted with the problem of identifying an alternative model that compensates for this inadequacy without introducing additional, even more erring suppositions. No method appears optimal for all conceivable situations. An exploratory, successive approximation-type approach can assist in understanding whether a method's assumptions (identified in a model scenario) are violated in actual studies of phylogeny.

Putting it all together: congruence and the iterative refinement of cladistic hypotheses

The previous sections lead up to the conclusion that there is a critical link between the process of individuating characters and the demands of an inferential standard on the evolutionary properties of those characters. If established observational practice would consistently produce information that violates the process assumptions of the preferred analytical tools, then the resulting inferences would be unreliable (to useless). On the other hand, a successful approach would emerge if the suppositions made by an algorithm for phylogenetic inference were addressed during the character-individuating stage. The most powerful inferential package would be obtained if there was some kind of mutual reinforcement among the two components, in the sense that the analytical tool offers a way to *test* whether its underlying assumptions are met by the available observational evidence. I will now argue that this is precisely what the cladistic practice of testing for congruence can achieve. The congruence test - a more conceptually wide-ranging implementation of the parsimony criterion - allows cladists to iteratively assess and increase the compliance of their homology statements with the process assumptions related to using parsimony as a phylogenetic inference standard.

Cladistic practice is sometimes characterized as an almost pedantic activity. Seemingly unambiguous observations are made, encoded in the character matrix according to simple procedures, organized at once with parsimony, and then without further analysis the results are submitted for publication. This is too easy. In reality, it is questionable whether any published cladograms are

identical to those originally obtained. In the course of a comprehensive project, cladists will often compute 10s or 100s of analyses for as many modified versions of an initial character matrix. Each analysis can vary in the number of characters and taxa included or excluded, and in the ways their properties are represented (encoded). This is more than just curiosity. What Hennig (1966) called *reciprocal illumination* is a sophisticated, empirical method aimed at improving the inference achievements of cladistic hypotheses. The practice of testing for congruence addresses virtually all of the aforementioned criteria for successful phylogenetic inference. It is in itself an implementation of (a cladistic notion of) parsimony. The testing for congruence contributes to refinements of homology statements, as well as to evaluations of the evolutionary properties inherent in the observed sets of characters. It can thus increase the precision and reliability of cladograms, explaining in part why parsimony is used to investigate phylogeny.

The initial results of an ongoing cladistic study rarely make it into print. There are many plausible reasons for refining them. For instance, the traditional criteria for assessing homology are not always unambiguous or infallible.[6] Conflicting interpretations of character state assignments may be analyzed in separate matrices, and optimized along the respective cladograms. The presence (or absence) of congruence with additional characters will likely strengthen (or weaken) one interpretation over its competitors. The original cladograms can also yield less resolution than expected. When confronted with multiple, equally parsimonious cladograms and a largely unresolved consensus, cladists tend to remain unconvinced that the results indicate actual concerted events of radiation. Instead, they will reconsider the encoded observations, and add characters or taxa to reduce the extent of ambiguity in the character matrix. The intention is to obtain a unique nested set of relationships among taxa, one that reflects phylogeny as accurately as possible.[7] In other situations, relevant (if somewhat extrinsic) information may contradict the structure of a particular cladogram. If an analysis of DNA sequence data places a spider within winged insects, then there was probably an error in assigning observations to taxa. This would be corrected upon investigation. More complex examples include theoretically implausible positions of clades or individual taxa, and unlikely optimizations of character states along cladograms. Certain results may posit the evolution of an old taxon on islands that originated more

[6]Rieppel (1988, pp. 44–48) presents various examples in animals where alternative statements of homology are possible. Each interpretation is contingent upon arbitrarily selecting one structure for positional reference. Moreover, 'ontogenetic repattering' can render the topographical criterion uninformative across more inclusive sets of taxa. Process-related ontogenetic criteria must be considered in such cases, though they can be ambiguous too. As Hall (1995, p. 25) reviews, 'there is sufficient lability in developmental mechanisms and sufficient hidden genetic variability underlying homologous features that no one-to-one relationship need exist between commonality of developmental/genetic bases and homology.'

[7]Another stated motivation for obtaining the most resolved cladogram is that this would optimally represent the information content, or taxonomic efficiency, of the character matrix (Farris 1983).

recently, or place a clade at the apex of a phylogeny when its paleontological record already indicates an earlier presence. A case of an implausible character state optimization would be the combination of small body size, nakedness, and endothermy in the hypothetical ancestor of arctic mammals. From time to time such propositions will make it into print. More commonly they are intercepted at the initial stages of a cladistic study.

From a philosophical stance, the congruence test may be understood as a way to assess the *projectibility* of particular homology statements, and of the resulting cladistic hypotheses. Contemporary software applications will compute a tree from a matrix and optimize the character states along its branches. This visual representation of parsimoniously structured observations provides information about the underlying homology propositions that was previously unavailable to the systematist. The exact congruence among various independent sources of evidence will increase the projectibility of each of them. Other (homoplasious) characters will be aligned with the overall topology *only* at less inclusive levels than first assumed. In the case of convergent properties in largely unrelated clades, the structural and positional similarities among them may be reanalyzed and reinterpreted, resulting in more precise statements of homology. Some observations may have entailed ambiguities from the start, and will conflict with the cladogram or confer no informative signal. As a consequence of this decrease in projectibility, they may be excluded from the last version of the matrix submitted for publication.

Some clarifications

At this point I am inserting various comments primarily for systematists. As circumscribed and practiced by Hennig (1966), the testing for congruence is inductive in nature.[8] It aims at producing iterative refinements of cladistic hypotheses. This is how parsimony considerations may influence the proposition of homology statements at rather elementary levels. In spite of claims to the contrary, such practice does not have to be viciously circular (Hull 1967).[9] Yet it is certainly a probabilistic approach. In a sense, the likelihood of each shared property as an indicator of phylogeny is evaluated in light of the presence (or absence) of congruence with other such attributes. The underlying causal assumption is that congruence would be *improbable*, unless a common origin existed.

Cladists themselves do not always stress the true nature and considerable impact of the congruence test in their publications. In some cases their

[8]Rieppel (2003) provides insights into Hennig's own conceptual account of phylogenetic systematics.

[9]The conclusions about homology and phylogeny are (typically) not entailed in the premises. A more appropriate metaphor than a circle would be a spiral of successively enriched and refined inferences. Later stages of analysis can affect the interpretations of earlier observations.

philosophical convictions would probably make an inductive, probabilistic method appear '*ad hoc*' and unjustified. Moreover, my account does not imply that all cladists use this iterative approach to the same extent. Individual researchers will have diverging attitudes as to what counts as relevant evidence, especially when it comes to more theoretical considerations of evolutionary projectibility. That said, if the congruence test really offers a way to produce more reliable inferences of phylogeny, then it would be curious if cladists did not take advantage of this method. And somewhat contradictory, considering that the predictive powers of cladograms (inferential reliability) are portrayed as one of the strengths of the cladistic approach.

One last clarification for systematists seems appropriate. The iterative practice of testing for congruence is aimed at reducing ambiguity and imprecision in a character matrix. This is not the same as eliminating all incongruent characters – a strategy implemented in so-called compatibility analyses. Homoplasy that is not just erroneous observation can be informative (homology at less inclusive levels), uninformative ('noise'), or misinformative (deceptive biases). Experience indicates that abundant yet informative homoplasy is as necessary as homology for reliable phylogenetic inference. In addition, considerable levels of uninformative homoplasy may be represented in a character matrix without affecting the overall conclusions (Wenzel and Siddall 1999). Conflicting evidence must be reanalyzed in a more comprehensive study. Critical reconsiderations of earlier homology statements can at times lead to less well supported, and less resolved cladograms. Thus, in spite of persistent attempts, there are epistemic limitations to the refinement of cladistic hypotheses.

Congruence and the exploration of molecular phylogenetic evidence

Previous examples have emphasized the practice of examining (individual) character state optimizations along cladograms. This method is well suited for analyzing phenotypic information, for example morphological characters. As described below, the testing for congruence can also lead to more reliable inferences based on DNA sequence data. Molecular systematists are no different from morphologists when it comes to using an iterative, parsimony-implementing approach to understand the nature of their observations. However, less time is spent to refine individual propositions of similarity (A = A, C = C, etc., but see comments on sequence alignment below). The main motivation is to explore the evolutionary properties of the molecular information *as a whole*.

The actual phylogenetic signal of (multiple, unaligned) DNA sequences is not easy to comprehend through visual examination. It is only through the algorithmic structuring with parsimony that cladists can evaluate this information, and compare it to other available data. In a sense, the moment after computing the cladogram represents the first time the data are 'seen.' One

could say that cladists learn most of what they know about molecular evolution by doing, where 'doing' means running parsimony analyses, and comparing the resulting cladograms. In direct analogy to morphological studies, the presence (or absence) of a congruent signal from several molecular loci translates into an increase (or decrease) in the projectibility of each of them. Mutually reinforcing cladistic hypotheses would be considered improbable unless indicative of common ancestry.

Iterative analytical practices that implement a sophisticated, wide-ranging notion of parsimony are abundant and diverse in molecular systematics. Here I can only focus on a select few, and summarize how they can address the requirements for a phylogenetically informative character matrix (see above). For instance, the process of aligning of multiple DNA sequences is analogous in principle to the stage of making initial assessments of similarity. A common problem is the interpretation of varying apparent insertions and deletions ('gaps') in the sequences. This relates to questions about the *precision* and *weight* of homology statements. The methods used to answer them typically attempt to maximize the congruence among all individual propositions (Mindell 1991). In some applications, the projectibility of each possible choice of alignment and gap placement is directly weighed against its contribution (or lack thereof) to the identification of a more parsimonious cladogram (Wheeler 2001). Parsimony analyses also function to detect and remove inadequate molecular evidence tied to events of paralogy and xenology (Doyle 1992). Unjustified assumptions about the *hierarchy* and evolutionary *independence* of the observed sequences are exposed in this way. Although parsimony imposes a nested structure on any kind of data set, the presence of an unresolved consensus among many equally parsimonious solutions may indicate tokogenetic (among-population) or reticulate relationships. Minimally, it suggests that the observed sequences are unable to resolve clades that evolved during parts or all of the covered time period.[10] Last not the least, parsimony assists in diagnosing the amount of *heterogeneity* and *transformation* represent in a character matrix. The method of optimizing character states provides a minimum estimate of evolutionary rates. It also provides information about the abundance and distribution of homoplasy along different branches of the cladogram. Invariable or excessively variable loci can be differentiated from those with adequate transformational properties. Instances of long-branch attraction may be identified by alternatively excluding one of the distinct branches (Pol and Siddall 2001). Other characters and taxa could be added to reduce the significance of this phenomenon.

[10]In time, the experiences acquired from such a trial-and-error approach 'teach' cladists about the potential of various molecular loci for solving a phylogenetic problem. Consider the research proposal: 'previous analyses using gene X were instrumental in resolving relationships within clade Y; therefore gene X is now selected to study the phylogeny of the closely related clade Z.' Implicit in this proposal is a parsimony-informed projectibility assessment about the properties of unobserved DNA sequences.

In summary, the validity of the evolutionary process assumptions implied in the use of parsimony is testable, through iterative practices that are in themselves implementations of the cladistic notions of parsimony and congruence. The reinforcing and self-correcting character of a powerful inferential method is thus obtained. I will return to this conclusion at the end and examine how it affects established philosophical views about parsimony and the cladistic approach.

Congruence and the entrenchment of a phylogenetically reliable terminology

Moving from the first towards the second major theme, I now focus on the workings of cladistic congruence over longer time periods. The linguistic aspects of systematics – the aim to provide a phylogenetic reference system for the biological disciplines – will receive more attention. The cladistic ideal is attained when an analysis produces not only reliable inferences of phylogeny, but also more precise statements of homology. The latter are natural characterizations of clades. As such they may function as important vehicles for formulating and testing questions in evolutionary research.

In addition to refining proportions of homology in the course of an individual study, the testing for congruence makes a contribution to phylogenetic reference over larger time scales. What Goodman (1954) calls *entrenchment* is the continuous embedding of certain predicates or theories within a linguistic tradition. It is a consequence of their successful records of actual past projections. Entrenchment occurs in cladistics when the referential extension of an earlier homology statement is reconfirmed by a later, more comprehensive analysis. Well entrenched phylogenetic terms should have the quality of corresponding to observations whose (parsimoniously inferred) congruence with other relevant evidence has been tested many times.

In the end, modern systematists do not accept 'spinnerets' as a predicate to individuate spiders merely to preserve a convention of the Linnean era. Rather, for centuries the presence of spinnerets have been weighed against an increasing amount of morphological and molecular attributes. This process has without exception led to the conclusion that they are indeed a (in this case unreversed) synapomorphy for the monophyletic clade of spiders. The inductive reliability of this conclusion is mirrored in the incorporation of 'spinnerets' in popular and scientific textbooks, in classes at schools and at research universities. To be sure, the use of this concept in everyday language might continue in spite of considerable referential imprecision. The perpetuation of terms in systematics, however, is contingent upon their correspondence to projectible, congruent evidence – which is where parsimony considerations come into play. Another example clarifies this point. Unlike most people, cladists cannot individuate reptiles in reference to unreversed ('spinneret-like') synapomorphies *unless* birds are subsumed under the concept of reptiles as well. Such counterintuitive practice illustrates how cladistic parsimony can affect the

entrenchment of projectible homology statements over time. Within the language of systematics, 'spinnerets' are indicators of relationship, and conducive to reference in ways that no terms corresponding to properties of reptiles while excluding birds could ever be. Only the former predicate has become entrenched during the continued process of testing for congruence.

It is possible to appreciate the effects of parsimony on the entrenchment of homology statements without having to think that there is an exact one-to-one correspondence among phylogenetic terms and natural phenomena. Predicates like 'triploid endosperm' or 'spinnerets' are exceptional in their precise referential extension to unique and unreversed synapomorphies in the living world. Most cladistic terms for homologies read more like this: 'five tepals connected,' or 'tibia with a spine-like projection.' Viewed in isolation, they could refer to a number of non-homologous states. They will remain ambiguous and less conducive to phylogenetic reference until the relevant taxonomic context has been specified. The contextuality of homology statements is even more apparent in DNA sequence data. By itself the predicate 'ACGT' is almost meaningless as an indicator of relationship. In turn, 'ACGT at position X of the molecular locus Y in taxon Z' can refer to a synapomorphy. Phylogenetic reference is like much of everyday language in this sense (Nunberg 1979). The long-term testing for congruence contributes to the entrenchment of homology statements. It cannot eliminate the context-specificity and linguistic redundancy of predicates.

A case of entrenchment in progress

The following is an example where the dynamics of entrenchment via congruence are not yet resolved. At this moment the referential extension of 'halteres' in insects remain controversial. Since the period of pre-Darwinian taxonomy, the term has been assigned to the modified (stalked and clubbed) *hind* wings of flies. However, recent cladistic analyses suggest a more inclusive meaning. According to these results, flies are most closely related to an enigmatic lineage called twisted-wing parasites (Whiting et al. 1997). The males of the latter have modified *fore* wings that appear to be similar in structure to the halteres of flies. For the two kinds of wings to be considered homologous, it is necessary to assume a past event of homeotic transformation, responsible for expressing the haltere-regulating molecular loci on a different body segment in each group. In other words, the exact referential extension of 'halteres' is partly contingent upon the acceptance of a theory about ontogenetic mechanisms. If developmental genetics accumulates evidence for the homeotic transformation hypothesis, then 'halteres' should refer not only to the wings of flies but also to those of twisted-wing parasites. Still, the latter interpretation remains less convincing on its own than it would be if supported by phylogenetic studies. More comprehensive parsimony analyses could for example indicate that flies and twisted-wing parasites are not closely related. In that case 'halteres' would

have evolved convergently. The term itself will have less referential precision (in the phylogenetic sense) than 'spinnerets,' even if the corresponding structures in each group are regulated by the same molecular loci.

Until now these issues have not been resolved. The situation is also complicated by the suspicion that the presumed monophyly of flies and twisted-wing parasites is an artifact of long-branch attraction (Huelsenbeck 1998). Regardless of the eventual outcome, it appears certain that the process of entrenchment of 'halteres' will be directly affected by phylogenetic analyses. Not long ago, it had been proposed that twisted-wing parasites are highly modified beetles, and their fore wings homologous to the hardened beetle elytra. But cladistic analyses have been unable to confirm this assignment. In light of this lack of congruence, the term 'elytra' continues to be used exclusively for the synapomorphic wings of beetles.

Evolutionary homeostasis, reference, and the limits of the cladistic approach

Although the present account is enthusiastic about the achievements of cladistic practice, it cannot offer a universal justification for using parsimony to infer phylogeny. This conclusion is already implicit in a probabilistic view that acknowledges the counterfactual significance of the phenomenon of long-branch attraction. I will now reassess the limitations of cladistics and direct more attention towards the strengths of maximum likelihood inference. The concept of evolutionary homeostasis will be useful in characterizing a basic difference among these competing approaches.

In the following it is necessary to separate the task of inferring phylogenetic relationships from that of establishing a system to individuate clades in reference to their evolutionary properties. Even though they are closely related in traditional systematics, it is possible to circumscribe a clade by listing its constituent members, or by pointing at a section of a tree. No organismal properties would have to be mentioned. Moreover, a particular method of inference could be more powerful or appropriate for one task than for the other. Evolution itself may render one of them more attainable. Because the conception of homeostatic property clusters (Boyd 1999) mixes an account of reference to natural kinds with the possibility of those kinds having 'evolving properties,' it is suitable for describing the difference among the two tasks.

As reviewed above, parsimony algorithms can run into problems when high evolutionary rates are combined with non-random inequalities in the abundance of homoplasy in unrelated lineages. Here I want to suggest that these limitations manifest themselves in a region within the universe of evolutionary processes that is largely incomprehensible using traditional approaches to reference and circumscription. Whenever parsimony as an inference standard computes an incorrect phylogeny, the threshold for verbal, property-centered characterizations of clades has probably been surpassed as well. Slow to moderate rates of evolution are required for such characterizations. Thus, any

inconsistency between parsimony and maximum likelihood inference concerns parts of the history of life where traditional reference is compromised. The latter appears to be less robust to the breakdown of evolutionary homeostasis than parsimony is in itself.

One can understand the difference in robustness by positing a scenario of linear saturation of DNA sequences with homoplasy. Parsimony will underestimate the evolutionary substitution rates in these situations. The margin of error could be very significant. However, so long as there remains an unambiguous signal in the character matrix, the error would not affect the capability of parsimony to infer the correct relationships. The same is not true for the task of referring to the individual character states that have succeeded each other along the way. Even if the overall topology has been reconstructed with accuracy, high rates of transformation could have made it impossible to describe all instances and structures that evolved along its branches. This may occur in analyses of rapidly evolving molecular loci. It is also a problem in morphological studies, whenever a set of structures has been modified throughout evolution to such an extent that the intermediate stages can no longer be recognized. The point is almost trivial, if one considers for example the impact of extinction on a complete phylogenetic account of morphological transformations. The ideal of precise reference to historical events is simply less robust to higher evolutionary rates.

Cladistic practice is well suited for verbal statements of homology. The latter are adequate if there is a largely unambiguous correspondence of a term to a causally sustained property which evolved at a specific phylogenetic juncture. Phylogenetic reference under the use of parsimony requires that evolution is not static. It succeeds when properties transform at moderate rates. As the rates increase and evolutionary homeostasis decreases, the inference of phylogenetic relationships may still be possible. The contribution that parsimony can make to a traditional reference system will be reduced. Often there will be multiple equally parsimonious optimizations of character states along the different branches of a cladogram. Should matters brake down further and introduce deceptive biases in the abundance of homoplasy, then parsimony will compute an unresolved or incorrect topology. The potential for proposing and refining traditional homology statements has been lost as well.

Probabilistic approaches

The impact of varying levels of homeostasis on phylogenetic reference is less significant if one adopts a numerical, explicitly probabilistic conception of similarity (Nielsen 2002; Sober 2002). This approach is used in maximum likelihood inference of molecular evidence. Now each character state along the cladogram is associated with a specific probability value. At low evolutionary rates, the probability of a particular state (A) should converge on '1,' whereas those of the alternative states (C, G, T) should be close to '0.' Such an

interpretation of a homology statement could be reconciled with its cladistic analogue.

The remarkable property of numerical circumscriptions is that they can accommodate large amounts of ambiguity without turning uninformative. As the rates of evolution increase and deceptive biases accumulate in the DNA sequences, the probabilities of individual character states along each branch will become similar but not identical. Taken to an extreme, a 'homology' for a clade may look like this: 'has state A with a probability of 0.6, *and also* state C with a probability of 0.3, *as well as* states G and T with probabilities of 0.08 and 0.02, respectively.' This example actually involves a simplification. The most appropriate way to represent the concept is through graphs for each state that depict its own distribution of probability values. In a related case, if the intent is to specify the optimization of a character state, this would read like: 'state A evolved in clade (X(YZ)) with a probability of 0.2, in clade (YZ) with a probability of 0.9, and in clade Z with a probability of 0.98.' From a probabilistic stance, this is a perfectly acceptable and informative circumscription of phylogenetic similarity.

Probabilistic approaches to homology and phylogenetic references are inherently more robust to ambiguity than verbal concepts. They are suitable in particular for conveying whatever information may be extracted from rapidly evolving character systems with a limited number of possible states. Yet the added resolution is difficult to translate into traditional diagnoses. Acknowledging this problem requires no sophisticated account of reference: whereas the imprecision of a predicate like 'tibia with a spine-like projection' is related to its contextuality, the aforementioned predicates are *structurally* ambiguous (Bach 1998). They have multiple plausible meanings in the same phylogenetic context. An analogy in everyday language would be the term 'English history teacher.'

My analysis does not imply that structural ambiguity cannot, or should not, be a characteristic of adequate phylogenetic reference. Some may argue that it is counterintuitive to think about such historically contingent, all-or-nothing phenomena as homology and synapomorphy in terms of partial probabilities. However, in certain situation this will be the only way to extract a signal from a character matrix. Systematists have already moved in this direction. So there remains another claim: the more structural ambiguities accumulate in statements of phylogenetic similarity, the less conducive these statements will be to a traditional reference system and to predictive classifications in systematics.

Species and clades are properly named and diagnosed in reference to homeostatic characteristics that are at the most contextually ambiguous. Visual displays of probability distributions can represent attributes that are even less stable in evolutionary space and time. Yet in spite of the relevance of their contents, they may be unable to play the aforementioned classificatory role. The reason for this seems to be a fundamental difference between the requirements for verbal reference and the ephemeral nature of certain evolutionary phenomena to which taxonomic predicates are supposed to correspond. One might say that evolution passed on too quickly to be captured in words.

Observing the threshold for traditional phylogenetic reference

The concept of evolutionary homeostasis can be used to characterize the point beyond which the cladistic notions of similarity and kind seem no longer appropriate. As outlined above, it is correct to interpret cladistic practice as a probabilistic approach to phylogenetic inference. The exact congruence among character states is considered improbable unless they had a common origin. It is also possible for cladists to deal with diagnostic ambiguity, for example by publishing alternative, equally parsimonious cladograms and character state optimizations. Clearly though, the probabilistic view of maximum likelihood inference adds another dimension to the notion of phylogenetic similarity. It is unclear how visual displays of probability distributions can have the same natural kind status within cladistics as conventional accounts of homology.

Boyd (1991, 1997, 1999) argues that the naturalness of natural kinds is manifested in their ability to support inferential practices established in particular scientific disciplines. This would suggest that (radically) probabilistic conceptions of similarity are not 'natural' with regards to the cladistic tradition. They make reliable contributions to a related but ultimately different tradition, often called statistical or molecular phylogenetics. It is a tendency within the latter research paradigm to consider the activities of naming and diagnosing to be less relevant than the objective of accurate phylogenetic inference. The predictive diagnostic powers of a natural system are often sacrificed to achieve this end. There may be a valid reason for this. Certain evolutionary phenomena investigated by molecular phylogeneticists are difficult – if not impossible – to translate into traditional circumscriptions and classifications. In some cases a phylogeny may only be retrievable using super-parsimonious algorithms and non-verbal concepts of relatedness. Such phenomena mark a point beyond which explanations for success cannot be offered under cladistic conceptions of similarity and kind.

In line with a naturalistic approach to analyzing science, I propose that the threshold for one or the other representation of homology in systematics is only *a posteriori* specifiable. Moreover, it should vary among character systems and taxa, and with the inferential resolution one wants to obtain. A way ahead is to examine actual practice and see where the relevance of diagnostic homology statements begins to disappear. For instance, it is uncommon in comprehensive molecular studies to individuate clades in reference to primary DNA attributes. Most of the time systematists will not report partial sequences like 'ACGT' as synapomorphies of inferred clades. Instead, they tend to make assertions about monophyly in light of support values, or other assessments of clade stability. In a sense, these concepts take the place of conventional homology statements which would be reliable only at less inclusive levels. Shorter, rapidly evolving DNA sequences appear to be most appropriate for diagnosing individuals, populations, and perhaps species; rarely for clades.

It is remarkable, however, that molecular systematists assign proper names to DNA primers. These are interspecifically conserved sequences with lengths of 15–30 bases. 'The DNA of clade X can be amplified with primer ITS3' is a homology assessment of sorts. Sufficient levels of evolutionary homeostasis also enable systematists to name many other phenomena at the molecular level, including secondary DNA structures and genome-level transformations. Often the latter have evolutionary properties similar to those of phenotypic characters like spinnerets. They can play a similar role in the elaboration of a phylogenetic reference system. Various nomenclatural practices in systematics are thus comprehensible if one adopts a naturalistic conception of homeostatic property cluster natural kinds.

Conclusions

I want to return to a question that is central to this explanatory account and runs throughout the paper. How is it that cladists can use parsimony successfully to infer phylogeny and to refine a natural taxonomy for the products of evolution? My intent was to demonstrate that answering this question requires more than an analysis of the inference capabilities of parsimony 'under almost any conditions.' A largely untackled problem is the ability of systematists to produce an initial set of observations with relevance to phylogeny. As soon as such propositions are available, the iterative method of testing for congruence offers an additional, sophisticated way to assess and increase their evidential status.

There are various projects in philosophy (e.g. Brady 1983; Sober 1988) and in systematics (e.g. Farris 1983; Steel and Penny 2000) arguing for or against particular methods to infer phylogeny. What they have in common is a tendency, though never absolute, to regard propositions of similarity as a 'necessary pre-condition.' This allows them to evaluate the performance of a method from a fairly removed perspective (supposed philosophical *a priori*, stipulated evolutionary process models, and so on).

Meanwhile, practicing systematists rarely (if ever) treat the challenge of producing relevant evidence as something more trivial than the algorithmic structuring of that evidence. In fact they might use the latter – only to aim at the former. Those who place too little emphasis on the effects of parsimony on the processes of character individuation and refinement may be unable to provide the kinds of answers systematists are interested in. For instance, even if at some point advocates of model-centered approaches succeed at specifying all conditions under which parsimony infers accurate phylogenies, they would still fail to capture an essential contribution of parsimony to everyday systematic practice.

A more complete account would emerge through consideration of a potentially positive inferential feedback loop. It appears that reliable phylogenetic inferences with parsimony *require* some level of evolutionary

homeostasis. The parsimony-implementing congruence test, in turn, can be exploited by cladists to selectively *achieve* observations reflecting this phenomenon.

It is hardly necessary to clarify that the present account does not aspire to *a priori* legitimize the use of parsimony in systematics. Several of the explanatory structures presented above apply to competing approaches as well. It is not as if pre-Darwinian taxonomists or likelihoodists (have) use(d) inference standards that have no relation whatsoever to parsimony.

Hennig (1966) proposed that there is a basic correspondence between nested summaries of character state transformations and many causally sustained relationships in nature. This assumption is known to be violated in some cases (Mindell and Meyer 2001) where 'non-hierarchical' inference standards may be preferable. On the other hand, the very existence of a verbal phylogenetic language indicates how stable and uniquely ordered much of evolutionary history must have been.

By now it should also be clear that the cladistic notion of parsimony is not just a simple way to apply an algorithm or Ockham's Razor. There are many reasonably attractive ways to apply 'the parsimony criterion' in systematics. Of all those, cladists have selected one that produces nested sets of characters and taxa, while making abundant use of incongruent information (convergence and reversal). 'Parsimony' assists in evaluating and improving the projectibility of the resulting phylogenetic inferences. Evidence that is relevant to the problem complex – only for theoretical reasons – may also be conjoined 'parsimoniously.'

Boyd (1985) argues that whenever scientists appeal to such non-experimental standards as maximal explanatory power or parsimony, their preference really reflects implicit, non-deductive assessments of *theoretical plausibility*. In the cladistic tradition, parsimony is a very pervasive and theoretically sophisticated concept. This explains why cladists are reluctant to call it 'minimum evolution.' In actual practice, parsimony can be almost synonymous with the iterative method of testing for congruence.

Acknowledgments

This paper evolved from a presentation at the 22nd Annual Meeting of the Willi Hennig Society (New York Botanical Garden, 2003). John Ascher, Richard Boyd, Caroline Chaboo, Huyen Do, Maureen Kearney, Roberto Keller, Kelly Miller, Olivier Rieppel, Elliott Sober, and Quentin Wheeler have all contributed to my understanding of science in more ways than I can acknowledge. Comments by the editor and an anonymous referee have assisted me in the challenging task of making the subject more accessible to a philosophical audience. My systematic research has been supported in part by the National Science Foundation Doctoral Dissertation Improvement Grant DEB-0206093.

References

Bach K. 1998. Ambiguity. In: Craig E. (ed.), Routledge Encyclopedia of Philosophy. Routledge, London. Available at http://www.rep.routledge.com/article/U001/

Boyd R. 1985. Observations, explanatory power, and simplicity: toward a non-human account. In: Achinstein P. and Hannaway O. (eds), in Observation, Experiment, and Hypothesis in Modern Physical Science. MIT Press, Cambridge, pp. 47–94.

Boyd R. 1991. Realism, anti-foundationalism and the enthusiasm for natural kinds. Philos. Stud. 61: 127–148.

Boyd R. 1997. Kinds as the workmanship of men: realism, constructivism, and natural kinds. In: Nida-Rümelin J. (ed.), Rationalism, Realism, and Revision, Proceedings of the Third International Congress, Gesellschaft für Analytische Philosophie. de Gruyter Verlag, Berlin, pp. 52–89.

Boyd R. 1999. Homeostasis, species, and higher taxa. In: Wilson R.A. (ed.), Species: New Interdisciplinary Essays. Bradford Books, MIT Press, Cambridge, pp. 141–185.

Brady R.H. 1983. Parsimony, hierarchy, and biological implications. In: Platnick N.I. and Funk V.A. (eds), Advances in Cladistics, Vol. 2. Columbia University Press, New York, pp. 49–60.

Brady R.H. 1994. Explanation, description, and the meaning of transformation in taxonomic evidence. In: Scotland R.W., Siebert D.J. and Williams D.M. (eds), Models in Phylogeny Reconstruction, Systematics Association Special Volume 52. Clarendon Press. Oxford, pp. 11–29.

Brigandt I. 2002. Homology and the origin of correspondence. Biol. Philos. 17: 389–407.

Cleland C.E. 2002. Methodological and epistemic differences between historical science and experimental science. Philos. Sci. 69: 474–496.

Doyle J.J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17: 144–163.

Farris J.S. 1983. The logical basis of phylogenetic analysis. In: Platnick N.I. and Funk V.A. (eds), Advances in Cladistics, Vol. 2. Columbia University Press, New York, pp. 7–36.

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

Goodman N. 1954. Fact, Fiction, and Forecast. Harvard University Press, Cambridge.

Griffiths P.E. 1994. Cladistic classification and functional explanation. Philos. Sci. 61: 206–227.

Hall B.K. 1995. Homology and embryonic development. In: Hecht M.K., MacIntyre R.J. and Clegg M.T. (eds), Evolutionary Biology, Vol. 28. Plenum Press, New York, pp. 1–37.

Hennig W. 1966. Phylogenetic Systematics. University of Illinois Press, Urbana.

Huelsenbeck J.P. 1998. Systematic bias in phylogenetic analysis: Is the Strepsiptera problem solved? Syst. Biol. 47: 519–537.

Hull D.L. 1967. Certainty and circularity in evolutionary taxonomy. Evolution 21: 174–189.

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

Kluge A.G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool. 38: 7–25.

Lennox J.G. 1991. Between data and demonstration: the analytics and the *Historia Animalium*. In: Bowen A.C. (ed.), Science and Philosophy in Classical Greece. Garland, New York, pp. 261–294.

Maund B. 2003. Perception. McGill-Queen's University Press, Montreal.

Mindell D.P. 1991. Aligning DNA sequences: homology and phylogenetic weighting. In: Miyamoto M.M. and Cracraft J. (eds), Phylogenetic Analysis of DNA Sequences. Oxford University Press, London, pp. 73–89.

Mindell D.P. and Meyer A. 2001. Homology evolving. Trends Ecol. Evol. 16: 434–440.

Nielsen R. 2002. Mapping mutations on phylogenies. Syst. Biol. 51: 729–739.

Nunberg G. 1979. The non-uniqueness of semantic solutions: polysemy. Linguist. Philos. 3: 143–184.

Quine W.V.O. 1969. Ontological Relativity and Other Essays. Columbia University Press, New York.

Platnick N.I. 1982. Defining characters and evolutionary groups. Syst. Zool. 31: 282–284.
Pol D. and Siddall M.E. 2001. Biases in maximum likelihood and parsimony: a simulation approach to a 10-taxon case. Cladistics 17: 266–281.
Rieppel O. 1988. Fundamentals of Comparative Biology. Birkhäuser Verlag, Basel.
Rieppel O. 2003. Semaphoronts, cladograms, and the roots of total evidence. Biol. J. Linnean Soc. 80: 167–186.
Rosch E. and Lloyd B.B. (eds) 1978. Cognition and Categorization. Lawrence Erlbaum Associates, Hillsdale.
Rosenberg A. 1994. Instrumental Biology, or, The Disunity of Science. University of Chicago Press, Chicago.
Ruse M. 1979. Falsifiability, consilience, and systematics. Syst. Zool. 29: 530–536.
Sober E. 1988. Reconstructing the Past: Parsimony, Evolution, and Inference. Bradford Books, MIT Press, Cambridge.
Sober E. 2002. Reconstructing the character states of ancestors: a likelihood perspective on cladistic parsimony. Monist 85: 156–176.
Sober E. 2003. The Contest between Parsimony and Likelihood. Unpublished manuscript available at http://www.aimath.org/wwn/geombio/
Steel M.A. and Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol. Biol. Evol. 17: 839–850.
Wagner G.P. 2001. Characters, units and natural kinds: an introduction. In: Wagner G.P. (ed.), The Character Concept in Evolutionary Biology. Academic Press, San Diego, pp. 1–11.
Wenzel J.W. and Siddall M.E. 1999. Noise. Cladistics 15: 51–64.
Wheeler W.C. 2001. Homology and DNA sequence data. In: Wagner G.P. (ed.), Evolutionary Biology. Academic Press, San Diego, pp. 303–318.
Whiting M.F., Carpenter J.C., Wheeler Q.D. and Wheeler W.C. 1997. The Strepsiptera problem: phylogeny of the Holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46: 1–68.
Winsor M.P. 2003. Non-essentialist methods in pre-Darwinian taxonomy. Biol. Philos. 18: 387–400.