Blackwell Publishing

Cladistics 22 (2006) 369–377

Cladistics

www.blackwell-synergy.com

# Rejecting "the given" in systematics

Maureen Kearney[1,*] and Olivier Rieppel[2]

*Departments of [1]Zoology and [2]Geology, Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, IL 60605, USA*

Accepted 27 February 2006

**Abstract**

How morphology and systematics come together through morphological analysis, homology hypotheses and phylogenetic analysis is a topic of continuing debate. Some contemporary approaches reject biological evaluation of morphological characters and fall back on an atheoretical and putatively objective (but, in fact, phenetic) approach that defers to the test of congruence for homology assessment. We note persistent trends toward an uncritical empiricism (where evidence is believed to be immediately "given" in putatively theory-free observation) and instrumentalism (where hypotheses of primary homology become mere instruments with little or no empirical foundation for choosing among competing phylogenetic hypotheses). We suggest that this situation is partly a consequence of the fact that the test of congruence and the related concept of total evidence have been inappropriately tied to a Popperian philosophy in modern systematics. Total evidence is a classical principle of inductive inference and does not imply a deductive test of homology. The test of congruence *by itself* is based philosophically on a coherence theory of truth (coherentism in epistemology), which is unconcerned with empirical foundation. We therefore argue that coherence of character statements (congruence of characters) is a necessary, but not a sufficient, condition to support or refute hypotheses of homology or phylogenetic relationship. There should be at least some causal grounding for homology hypotheses beyond mere congruence. Such causal grounding may be achieved, for example, through empirical investigations of comparative anatomy, developmental biology, functional morphology and secondary structure.


Hennig's (1966) "Phylogenetic Systematics" was a milestone in systematics, especially notable for its insistence on monophyletic groups and grouping by synapomorphy. However, another crucial element of Hennigian phylogenetic systematics is its emphasis on initial character analysis as a necessary guide to homology. Hennig (1966) used a variety of criteria (such as detailed comparative morphological studies, topology, connectivity, ontogeny, functional anatomy, geological precedence in the fossil record, and ecology) to identify, analyze and polarize characters. Evaluation of character quality and utility were based on both theoretical justifications and empirical investigations (see also Hennig and Schlee, 1978). Although one may disagree with the use of any or all of these guidelines, it is instructive to note that homology was something to be comprehensively investigated prior to tree construction for Hennig, not solely the result of phylogenetic analysis.

Pheneticists believed it impossible to make such judgments about taxonomic characters objectively (Sokal and Sneath, 1963; Sneath and Sokal, 1973). Sneath and Sokal (1973, p. 87) emphasized that approaches to character data need not be based on detailed biological evaluation: "One way to deal with problems of homology is to ignore details of structure…" Numerical taxonomists advocated equal weighting of all characters, the use of large numbers of characters, quantitative character coding, and a "theory-free" approach to character identification. Phenetic grouping methods were effectively criticized as untenable by biologists (e.g., Farris, 1983). In addition, the inherent scientific and philosophical naiveté of phenetics in terms of atheoretical data collection was exposed as the "look, see, code, cluster" approach (Hull, 1994). Fundamentally, the phenetic approach to character data reduces charac-

*Corresponding author:
*E-mail address:* mkearney@fieldmuseum.org

ters to raw observations, and this uncritical empiricism is one factor that ultimately led to the method's demise. However, the overall philosophy does not seem to have been completely overcome in modern systematics, at least when it comes to morphological characters. We identify two principle arguments in support of a similar approach to character delimitation. The first approach paradoxically acknowledges the fact that no theory-free observation is possible, but rejects theoretical and empirical evaluations of characters in favor of a putatively rigorous method of testing (congruence). The second approach emphasizes our ignorance with respect to the causal correlates of phylogenetically informative characters, and seeks as unbiased an approach to character delineation as is humanly possible. Both approaches ultimately defer to the total evidence criterion as the sole method of testing homology.

For example, it has recently been argued that biological evaluation of characters (investigations of potential character correlation, for example) is ''irrelevant'', and that the only criterion for inclusion of a character in a phylogenetic analysis is the concept of a transformation event (Grant and Kluge, 2004, p. 23). As such, ''anything'' can be a potential character hypothesis so long as a rigorous method of testing these character hypotheses is used: ''There is no one operation for determining character states in this system—it can be *anything* that leads to the testable hypothesis of synapomorphy'' (Kluge, 2003a, p. 356; emphasis added). The method believed to test these ''anythings'' is that of congruence under parsimony, and total evidence analysis of an increasing number of characters is believed to increase the severity of that test.

And, in reference to the biological evaluation of characters by Naylor and Adams (2001) (potential developmental correlation and non-independence in this case), O'Leary et al. (2003) reject the notion that independent coding of serial homologs (molars in this case) is problematic. They compare such evaluations to the ''intuitive'' and ''authoritarian'' approaches of evolutionary taxonomists, which are understandably deemed unscientific. In more general reference to the empirical analysis and potential rejection of previously published characters by systematists, they note ''a disturbing trend towards data selectivity.'' Citing the total evidence principle, they advocate instead that phylogenetic studies must include all previously published character data in a global congruence test, this being the most objective and rigorous way to test characters and homology (O'Leary et al., 2003, p. 861). Thus, every published observation becomes incorporated into an ever-expanding data matrix in order to avoid ''data selectivity'' or ''data exclusionism.'' Similar thinking about homology evaluation under the total evidence criterion and the Popperian paradigm was also evident in Rieppel (1992) and Kearney (2003).

The concept of bringing as much evidence as possible to bear on phylogenetic problems (e.g., Kluge, 1989; Eernisse and Kluge, 1993) is obviously a valuable one. However, as cited above, some contemporary systematists interpret the meaning of total evidence as the requirement of giving the phylogenetic tree complete logical priority over critical comparative studies of character data—from such a viewpoint, it is only the tree, not empirical character evaluation, that can inform us about homology and what a legitimate character might be (Härlin, 1999). This ultimately reduces to the contention that knowledge of homology and phylogeny can be gleaned from the pure coherence of theory-free observation reports on the ''immediately given''—whereas such knowledge claims have long been exposed by philosophers of science as impossible (Hanson, 1961; Hempel, 1965; Popper, 1992). As Ruse (1988, p. 60) points out: ''As soon as one starts breaking organisms into parts, one must bring in theory … Take two bears, one white and one brown. Do they differ in one feature, or does one take each hair separately … The point is whether someone who explicitly eschews the theory has the right to combine all the hairs into one feature.'' In our view, total evidence and the test of congruence *per se* are highly relevant to phylogenetic inference, but notions that the total evidence criterion obviates the necessity of character analysis, that any raw observation (or even a published descriptive statement) can be a character, or that congruence under parsimony (or any other optimality criterion) can be the sole arbiter of homology, are flawed.

## The philosophy of total evidence versus the philosophy of Karl Popper

The ''total evidence'' school of thought argues that the best phylogenetic hypothesis is the one that is based on the largest number of congruent data points. The test invoked in the context of the total evidence approach is the ''test of congruence'', which was originally introduced by Patterson (1982), but not in a Popperian context. Indeed, issues of congruence, covariance or consistency are not so much part of Popper's falsificationism (although the hypothesis/theory and the test statement must of course cohere: Audi, 2003), but rather play a role in a program dubbed ''coherentism'' by epistemologists (e.g., Kirkham, 2001; Lynch, 2001), where the coherence (non-contradiction) of statements on the evidence condition is considered both a necessary and sufficient condition for the acceptance or rejection of a hypothesis/theory (Blanshard, 2001).

Total evidence is a classical principle of inductive inference (see also Fitzhugh, 1997; Lecointre and Deleporte, 2004) that was appealed to by the empiricist philosopher Rudolf Carnap (see Carnap, 1997a,b, for an accessible discussion). The total evidence for Carnap (1997b, p. 972; emphasis added) comprised the "*total observational knowledge*" available to a person at the time of inference-making. As pointed out by Lecointre and Deleporte (2004, p. 102), this total evidence also includes a person's "background knowledge", which introduces the important distinction of "relevant" as opposed to "irrelevant" evidence. When introducing the principle of total evidence in systematics, Kluge (1989) cited the pertinent philosophical literature, i.e., Carnap (1950) and Hempel (1965); subsequent contributions advocating total evidence do so in the context of a purportedly Popperian philosophy and approach to systematics (e.g., Kluge, 1997, 1998, 2001a,b 2003b, 2004; Siddall and Kluge, 1997; Grant and Kluge, 2004). There are no grounds, however, for appealing to Popper's falsificationism in support of the principle of total evidence. Popper was explicitly not interested in the ampliative (successive and mutually reinforcing) accrual of evidence in support of a hypothesis/theory (Lipton, 2004; see also Fitzhugh, 1997; Rieppel, 2003a). In contrast to Carnap's "degree of confirmation", Popper's "degree of corroboration" is not a measure of increasing support that would make promises as to the future performance of a theory, but merely a historical account (e.g., Popper, 1989, p. 402), i.e., a record of past successes of a theory. In an elegant analogy, Godfrey-Smith (2003, p. 68) compared "degree of confirmation" with a "letter of recommendation" and "degree of corroboration" to an "academic transcript." The academic transcript only says something about the past performance of a student, while the letter of recommendation makes predictions about the future performance of a student that are based on her past performance. According to Popper, it is impossible to conclude from the fact that a theory passed a test in the past that the same theory will pass a *different* test in the future. In fact, Popperian methodology requires that a corroborated theory should be subjected to the severest kind of test possible, given current background knowledge, in an attempt to falsify it.

For Carnap, the "degree of confirmation" increases with an increasing number of positive instantiations of that theory's implications. For his "inductive logic" it did not matter whether evidence in support of a hypothesis was gathered before or after the formulation of that hypothesis (the order in which evidence is acquired is epistemologically irrelevant); all that matters is the "total set of statements available for assessing *h* at the end of the evaluation period" (Kitcher, 1993, p. 247). (The same philosophical attitude underlies the idea that all previously published data must be included in a total evidence analysis in systematics.) In contrast, for Popper, the ability of a hypothesis to accommodate statements that are already accepted is irrelevant to its status, for "it is too easy to make up a hypothesis that will fit antecedently known statements" (Kitcher, 1993, p. 246). What mattered for Popper are novel predictions generated by a hypothesis, and the riskier (in light of current background knowledge), the better. This is also the reason why a theory may fail the next round of testing, no matter how high its previous degree of corroboration.

Indeed, the only time Popper discussed the concept of total evidence is when he formulated his own *positive* solution to the problem of confirmation of scientific theories, a solution that differed from Carnap's. In contrast to Carnap, Popper's (1997, p. 221f) definition of degree of confirmation requires the *subdivision* of the total evidence into the old evidence relegated to background knowledge, and the new evidence that is critical to the confirmation of a theory (see Rieppel, 2003a; Lecointre and Deleporte, 2004). Such subdivision of the total evidence is certainly not what is meant by the use of "total evidence" today in systematics.

What, then, are the factors that have resulted in the linkage of "total evidence" and the "test of congruence" with Popperian falsificationism in systematics? We suggest that, to some degree, systematics continues to replay the famous "protocol-sentence debate" that earlier split the logical empiricists into coherentist and foundationalist factions (Hung, 1992; Carnap, 1997a, p. 57; Friedman, 1999). Protocol sentences were claimed to communicate observations, and hence were meant to relate to the "empirical base" of natural science. As character statements are (presumably) at some point related to observation, it is worth revisiting the "protocol sentence debate" in the light of current debates on the nature of characters and their testing.

## Rejecting the "given"

Numerical taxonomists (Sneath and Sokal, 1973, p. 17; italics in the original) stressed the "*empirical approach*" in taxonomy, where the "main emphasis" is on "firm observation and not upon "phylogenetic assumptions." In that context, Sneath and Sokal (1973, p. 17) cited P. W. Bridgeman's operationalism, a philosophy that sought to translate all theoretical terms of science into observational terms, such that theoretical concepts would become rigorously testable "by observation and experiment" (Sneath and Sokal, 1973, p. 17; see also Hull, 1968). This program could succeed only if there was a theory-free "observation language." Whether or not such a language exists was the central issue of the "protocol sentence debate." [Proponents of a total evidence approach in systematics correctly argue that no such theory-free "observation language" exists (Kluge, 2003a), yet at the same time admit *any* observation report of characters and states into the total evidence under evaluation and disallow empirical rejection of the same.]

The "protocol-sentence debate" (e.g., Oberdan, 1993) centered on the possibility, or lack thereof, of unmediated observation, i.e., of a direct acquaintance with what is immediately "given" in sensory experience that would then result in intersubjective and indubitable observation reports. Early versions of logical empiricism such as Bertrand Russell's "logical atomism" (Stroll, 2001) insisted on the immediate acquaintance of the subject with the "given" that is supposedly obtained through theory-free observation, and that would be communicable though "atomic propositions" such as "*this is white*" (the speaker pointing at a white table for example). Hennig (1950, 1966) himself turned to the positivist philosopher Theodor Ziehen for an analysis of the "given" (for discussion see Rieppel, 2003b). The problem with an appeal to the "given", one that Hennig (1950, 1966) was acutely aware of ("the participation of human activity in the process of perception … must make us suspicious": Hennig, 1966, p. 11), is that it does not explain how "private" (subjective) and fleeting perceptual experiences can provide the basis for intersubjective communication, nor for lasting empirical science. The positivist solution to this problem was the introduction of the verificationist criterion of meaning (e.g., Ayer, 1952; Schlick, 1959a). The "verification principle" held that the meaning of a statement was given by the specification of the conditions under which a statement could (at least potentially) be recognized as either true or false (the "verification principle" is really a principle of testability: Godfrey-Smith, 2003, p. 27). On that account, however, observation *reports* become observation *statements* that have propositional content, i.e., they can be right or wrong. This goes beyond Russell's logical atomism, where "atomic propositions" would not allow for error (Stroll, 2001). The introduction of the "verifiability criterion" thus broke the immediacy of acquaintance with the "given" that was requested by Russell. However, a problem ensued due to the fact that the explication of an observation statement, of the words it is composed of, and the evaluation of its truth value (i.e., in a discussion as to whether it would be true or false under certain conditions) will require other statements whose explication requires yet other statements, and so forth *ad infinitum*. Because the explication of observation statements will eventually run out of words, some empiricists proposed to resolve this infinite regress by an appeal to ostensive definitions (e.g., Schlick, 1959b). Such definitions establish the meaning of a term through ostensive indication of its referent, achieved by pointing at a certain observable particular (located in a certain space–time region) to which the term is to refer (e.g., "this table here and now is white"). Hoping, like Russell before him, to secure a firm foundation of knowledge, Moritz Schlick, leader of the foundationalist faction of the Vienna Circle, used ostension to establish the basis for what he called indubitable *Konstatierungen*. These would be rendered as present-tense indexical sentences such as "this now green". However, in their present-tense indexical form, such *Konstatierungen* cannot provide lasting foundations for empirical sciences; they can at best only provide justifications here and now for some given observation statement (Oberdan, 1993, p. 52).

We proposed above a persistent tendency in modern systematics to reduce character statements to immediate, putatively theory-free observations (the "given"). This is reminiscent of the general strategy pursued by those empiricists who wanted to ultimately base all science on pure observational reports as with Schlick's "*Konstatierungen*". Carnap (1997a, p. 57) specified in his autobiography, "we assumed [under the influence of the early Wittgenstein] that there was a certain rock bottom of knowledge, the knowledge of the immediately given, which was indubitable." However, some of the early empiricists, notably those that came to constitute the coherentist wing of the Circle around Carnap and Neurath, became fully aware of the unavailability of such rock bottom knowledge, following the lead of Popper in that respect (Carnap, 1997a, p. 57). Indeed, Popper's earliest claim to fame was his insight that all observation statements were, so to speak, "soaked in theory" (Popper, 1989, p. 387): "Each time when we take a reading from an instrument we rely on the hypotheses of geometrical optics, on the theory of solid bodies, on the correctness of Euclidean Geometry in small space, on the hypothesis of the existence of things, and innumerable other hypotheses" (Popper, 1979, p. 391). Carnap (1997a, p. 31) acknowledged that Popper "positively influenced" his thinking, most importantly through his "views on protocol sentences" (Carnap, 1997a, p. 32; "basic statements" in Popper, 1992). Popper emphasized that there cannot be any absolute, indubitable protocol sentences, an acknowledgment that introduced an element of conventionalism into his philosophy of science: "basic statements" can only be *accepted* as such by the scientific community.

The recognition that there cannot be a rock bottom of knowledge has potentially disastrous consequences for foundationalism—in contrast to coherentism, foundationalism seeks to bridge the gap between words and things, traditionally by the establishment of a correspondence relation as was articulated by Popper (1973, p. 46): "*A statement is true if and only if it corresponds to the facts*." This claim implies the methodological idea that the truth of a statement is established by comparing it, or its observational consequences, with facts. However, the rejection of a rock bottom of knowledge implies the acknowledgment that the gap between words and things can never be completely closed, that every statement (including protocol sentences) can be wrong and may become the subject of potential revision and/or correction at any time. Such fallibility pushed some members of the Circle towards coherentism (Neurath, cited in Oberdan, 1993, p. 32; see also Carnap, 1997a, p. 57). As is well known, Popper (1992) rejected a confirmationist analysis of science instead, and adopted a highly skeptical, falsificationist position.

Coherentism seeks only maximal coherence of the system of sentences for science, including protocol sentences or observation statements. The foundationalists of the Circle sought instead some perceptional touchstone for empirical sciences, described in terms of protocol sentences or observation reports that were epistemologically privileged in that they would "hook up" to the empirical world. In his discussion of what he called "basic statements", Popper (1992) contended that no such epistemically privileged statements exist. Instead, a "basic statement" simply becomes one through its (provisional) acceptance by the scientific community (but was still at least supposed to report on an observable state of affairs). However, as there cannot be any theory-free observation statements, the coherentists in the Vienna Circle went beyond Popper with their conclusion that *any* statement (not just one that relates to an observable state of affairs) can be accepted as a "basic statement", i.e., as a stopping point of inquiry into the world. From this perspective, all that matters for the acceptance of a statement is its coherence with other, antecedently accepted statements of the same domain of scientific discourse. Neurath followed what he identified as a Popperian tendency to "sideline protocol sentences" (cited in Oberdan, 1993, p. 36) and concluded that, although an empirical ("synthetic") statement is characterized by the fact that one can derive an observation sentence from it, one does not need to "descend all the way to the observation sentences" in the test of an empirical ("synthetic") statement. Instead, one can stop at any other statements (of the relevant domain of discourse) and take those for "protocol statements" (Oberdan, 1993, p. 36). This, of course, carried the threat of an all-out coherentism that completely abandons the perceptual base of empirical knowledge. However, although Neurath did indeed adopt a coherentist account of scientific knowledge, he did so not for reasons of metaphysics, but purely for pragmatic ones. He considered coherence of statements, both observational and theoretical, as a necessary condition for science, but *not also a sufficient condition*: "When Neurath characterized his methodology in terms of a comparison of a proposed sentence with antecedently accepted ones, he implicitly presumed that the sentence under consideration was responsibly asserted by a properly conditioned speaker, and not just randomly chosen" (Oberdan, 1993, p. 46). For Carnap, the coherence of statements on the evidence condition was similarly a necessary condition for natural science to proceed, but again not a sufficient condition: the coherent statements must still be confronted with observation (Oberdan, 1993). (Both empiricist foundationalism and coherentism were embedded in a phenomenological understanding of perception.) It is in this very same sense that we consider the coherence of character statements a necessary, *but not a sufficient*, condition for analysis of homology and phylogeny.

## Necessary but not sufficient

Epistemologists have analyzed both coherentism and foundationalism for their strengths and shortcomings (e.g., Leplin, 1997); in fact both are necessary and neither can stand on its own (Haack, 1998, 2000). The major problem of foundationalism concerns the justification of "basic statements" relative to observable states of affairs—coherentists will correctly point out that such justification can never be absolute. On the other hand, foundationalists will correctly point out that the mere coherence of any statements relative to one another will not suffice as a test that could potentially confirm or disconfirm a "basic statement" relative to an observable state of affairs.

The difference between foundationalist and coherentist philosophies seems relevant to some current debates in systematics. In spite of supporting an ostensively defined language of systematics (e.g., Kluge, 2003a) "observation theory" is rejected and simple coherence of character statements (congruence of characters) is turned to instead. Such a "coherentist systematics" considers the coherence of statements on the evidence condition as the necessary and sufficient condition for the acceptance or rejection of a homology hypothesis, rejecting the need for empirical and causal grounding for such statements. Classical points of criticism against coherentism apply to such an approach in systematics—for example, the lack of empirical grounding for character statements makes phylogenetic results vulnerable to analytical manipulation. The stance against biological evaluation of characters, or against any criteria for homology hypotheses, can result in a serious underdetermination of phylogenetic hypotheses (see also Richards, 2002, 2003). Such indeterminacy of character statements allows other researchers to exploit the system and potentially immunize their favored phylogenetic hypotheses. Through definition and re-definition, virtually any character statement (certainly of morphological characters) can be made to cohere with any set of other such statements, and through splitting or lumping of the number of character statements, the same can be achieved. This is particularly true if "anything" can be a character on the sole condition of its coherence with other characters relative to a hierarchy.

Coherentism, a path from which even Carnap (1997a) partially retracted, appears on the surface to allow circumvention of the problem of how to close the (logical) gap between words and things. According to Popper (1992), all statements, including protocol sentences, can potentially be false and must remain open to revision. There can never be any certainty of *empirical* knowledge. As a consequence, Popper abandoned the search for evidence in support of a theory, and turned to its potential falsification instead. However, the price to pay for this turn was the need to find *support* for a counterexample to the theory under test (Popper, 1989, p. 240), i.e., support for a statement that reports on an observable state of affairs that potentially falsifies the theory under test. Popper thought (erroneously: Lipton, 2004) that he could ignore the context of the origin of a theory. He asked that theories, or hypotheses, be simply posited and then put to test. This seems to correspond to the notion that ''anything'' can be posited as a character, as long as it can be put to test. *But the test of congruence is not a Popperian test*.

For Popper (1992), a theory is empirically testable only if it has deductive consequences that are forbidden (i.e., that are negated observation statements). If such consequences obtain, they would falsify the theory from which they were deduced. Being low-level hypotheses, character statements do have propositional content in the sense that they can be right or wrong but, given the radical contingency of the evolutionary process, there cannot be a deductive link between a hypothesis of phylogenetic relationships and character distributions on a tree (Sober, 1988); if there was, we would not need the test of congruence. In contrast, the ''test of congruence'' measures the degree of character congruence on which we base our judgments as to how well a phylogenetic hypothesis is supported (or, in the falsificationist mode, which of the alternative hypotheses of relationships is the ''least falsified''; Kluge, 1997). If observation is theory-laden, as Popper showed, and if that theory-ladenness varies with the individual worker's background knowledge and dispositions, then it might seem better to put no constraints on the evidentiary condition that bears on phylogeny reconstruction other than mere congruence (coherence of character statements)—that, in a nutshell, is the coherentist systematist's point of view. Put another way, if the conceptual (logical) gap between words and things cannot at least in part be bridged, coherentism may seem to be the only way to make sense of the world. However, the mere coherence of character statements in itself does not root those in the causal structure of the extralinguistic world. Coherence (non-contradiction) is a logical, or inferential, relation between character statements (propositions); those statements need to be linked to the organisms under investigation in a relevant way because the work that congruence will do is only as good as are the characters entered into the analysis. There is no question that coherence of character statements relative to a hierarchy is a *necessary* condition of phylogeny reconstruction, but it is not also a *sufficient* condition.

A coherentist systematics rests on the belief that the severity of test increases exclusively with an increasing *number* of characters used in phylogenetic analysis, no matter what exactly the nature of those characters is (they can be ''anything'' so long as they are ''logically independent''). One reason to believe so is that the higher the number of characters used, the less likely is their congruence. This might be true if each character corresponded to some bit of information that could be empirically grasped by every working systematist, and that would be fully independent from all other such bits of information. This, however, is plainly not the case, for biological as well as epistemological reasons. Biological reasons include the potential interdependence of characters (sometimes related to the developmental or functional correlation of characters). Epistemological reasons include the distinction of ''irrelevant'' versus ''relevant'' evidence (Naylor and Adams, 2003), as it results from the relation of evidence to background knowledge (Lecointre and Deleporte, 2004). In the domain of systematics, character statements are nothing but low-level theories, i.e., hypotheses of putative homology. Not ''everything'' or ''anything'' passes as a relevant character statement in the light of prior practice that has, after all, previously rejected the phenetic punch-card approach to character delineation. Logical empiricists of the foundationalist camp overestimated the powers of observation, given that the partnering of observations with background knowledge is itself a relation of coherence, and yet it remains ''the evidence of the senses [that] ultimately anchors our theories in the world; and it is a real constraint'' (Haack, 2003, p. 125).

## ''Look → See → Code''

A common notion in systematics is that objectivity requires the rejection of theoretical dependence. However, it is widely acknowledged by most scientists and philosophers—and is indeed the essence of the Duhem–Quine thesis—that the relation of hypotheses to empirical evidence is intimately dependent on theory in all fields of science (e.g., Godfrey-Smith, 2003; Lipton, 2004). It is theory that provides criteria for identifying justifiable grounds of inference, the appropriate areas in which to search for empirical tests, and the appropriate methodological designs for doing so. For example, to bring the insights of developmental biology, functional anatomy, detailed comparative morphology, and other evolutionary considerations to bear on character delineation and on the issue of character interdependence is to bring theory to bear on similarity relations (Rieppel and Kearney, 2002). To dismiss these areas of biological background knowledge does not escape theoretical process assumptions, but rather adopts a theoretical and philosophical stand that approaches the character issue from an instrumentalist and operationalist perspective.

Rieppel and Kearney (2002) argued that character conceptualization need not consist of primitive or abstract notions of similarity [although Kluge (2003a, p. 357) correctly criticized that argument for being cast in a falsificationist framework]. At the same time, it was acknowledged that criteria of character identification and re-identification remain necessary for empirical scientists to conduct comparative phylogenetic work. Traditionally, operational criteria such as structural details, topological correspondence and connectivity (Remane, 1952) have more or less successfully been used to help make the common historical origin of homologs empirically accessible, even in face of the fact that topological relationships can themselves evolve. Further, topological correspondence and connectivity are criteria used by all systematists today, whether implicitly or explicitly (Rieppel, 1994). It is assumed that this is so, not because of any arbitrary notion of similarity, nor because of a merely conventional use of topology and connectivity in the search for homology, but because these guides are at least approximately and defeasibly aligned with causal evolutionary and developmental processes (Rieppel and Kearney, 2002). In other words, traditional operational criteria of homology have been successfully deployed in phylogenetics presumably because they are, to some degree, well-aligned with the causal processes that relate to the evolutionary transformation of morphology. Such well-aligned criteria are what allow transcendance of "primitive" (unanalyzed) similarity (i.e., the outermost ear ossicle of a mammal and the lower jaw of a shark are not phenotypically similar but they share similar topological relations; such guidelines have arguably led to the successful discovery of homology, whereas "primitive" similarity could not).

Whereas Rieppel and Kearney (2002) suggested that character hypotheses should be as theoretically and causally grounded as possible (with the proviso, of course, that empirical science can go wrong), Grant and Kluge (2004, p. 26) argue that this cannot be so: "What matters in individuating character-states … is not the structural, developmental, or functional independence of a part, but its historical/transformational independence". This is a conceptually attractive point, but it is not empirically accessible—systematists do not observe historical and transformational independence or non-independence in their empirical work and thus cannot use such concepts in delineating characters. Even within the framework of research cycles (Kluge, 1998), there exist empirical criteria for (re)examination of characters.

One example in vertebrate phylogenetics that illustrates the tenaciousness of the problem is the snake origins debate (discussed in Rieppel and Kearney, 2001, 2002). For example, how many independent characters should be coded for hind limbs versus features of the skull in fossil snakes is critical to alternative phylogenetic outcomes. And, if anything can be a character, the problem only worsens. Controversy about tetrapod origins also replays this character debate to some extent (Schultze and Trueb, 1991). Coding of certain dental features, a matter of contention in studies of whale origins (O'Leary and Geisler, 1999; Naylor and Adams, 2001; O'Leary et al., 2003), provides another good example of the problem. Prior knowledge and practice suggests that "teeth should be considered not only as individual units but as part of a system … [morphogenetic] fields can often be classified as corresponding to the incisor, canine, and premolar-molar regions of the dentition and each of these regions appears to have some degree of evolutionary independence" (Scott and Symons, 1977, p. 82). On that developmental basis it appears unproblematic to code characters for canines, incisors, premolars and molars separately. But, as Osborne (1978, p. 196) emphasized: "teeth within a class are, like hairs or leaves, serially homologous structures and … it would be … meaningless to look for (historical) homologies." O'Leary et al. (2003) reject the theory of developmental correlation of molar features and rely solely on total evidence and the testing procedure (congruence/parsimony) to assess homology. But if there are good empirical reasons to believe that certain characters are developmentally correlated, then the character statements denoting those characters will not be expected to contradict each other, i.e., they will not be expected to fail the test of congruence. They will be expected to cohere, and, if enough of them are coded, they risk skewing the analysis in a certain direction. O'Leary et al. (2003) correctly point out [in response to the criticism of Naylor and Adams (2001) regarding non-independence of molar characters] that developmental correlation of molar characters is not known beyond doubt for all mammalian species—but neither will anything be in the realm of empirical science. We wonder whether that bar of knowledge is set too high.

Coding suites of developmentally correlated characters as independent entries in a data matrix does not avoid theoretical baggage, nor successfully escape criteria of homology. Instead, such coding takes a theoretical stand, which is to *a priori* reject the hypothesis of developmental correlation. Such a rejection also rejects previous comprehensive empirical studies that demonstrate causal correlation. And, even if it is believed that evolutionary process theories can be avoided by taking the theoretical stand of dismissing them, prior practices can still not be avoided. It is, after all, prior practice that informs systematists about the difference between premolars and molars, about the nature of molar cusps, and about the comparability of these cusps across taxa. In contrast, it may be argued more successfully that developmental correlation of characters can at least to some degree be experientially and causally grounded.

## Conclusions

The de-emphasis (and even abandonment—e.g., Wheeler, 1996) of character analysis and primary homology hypotheses represents a familiar shift from the theoretical and empirical foundations of comparative biology to an instrumentalist–operationalist approach to systematics. With the latter approach, only the *number* of shared character statements, or the semantic definition and redefinition of the same, determines homology and topology. Characters are reduced to mere instruments that allow choosing among the finite number of possible alternative hypotheses of relationship for any given number of terminal taxa according to some optimality criterion (e.g., parsimony). However, at least since Darwin, systematics is meant to be an empirical science, and biological classification (or "systematization"—Griffiths, 1974) is meant to capture the relevant part of the causal structure of the world. Coherentism alone seems too weak for biological systematics, which provides the important foundation for many fields of evolutionary biology.

Phylogenetic character data need to be empirically and causally grounded, at least to some degree, rather than merely consisting of "unprocessed observations". If there is no theory-free observation (and there is not), then there is also no theory-free character conceptualization or coding. Even to apply a punch-card (Sneath and Sokal, 1973, fig. 3-1) to the automatic scanning of characters in a purely operational/algorithmic approach still presupposes theory, namely the theory that morphological homologs can be discovered by ignoring anatomical complexity. The positivist empiricism inherent in the phenetic program, aimed at developing methods that are "objective, explicit, and repeatable" (Sneath and Sokal, 1973, p. xii) remains a myth. The solution to the difficult problem of character conceptualization and coding cannot be to try to avoid theory in observation or to accept any raw observation or published character statement as a phylogenetic character. A partial solution may be to recognize that congruence of characters, although a necessary condition for phylogeny reconstruction, is not a sufficient condition.

It is possible to explore the potential for character interdependence both empirically and heuristically and we believe these efforts, thought difficult and time-consuming, are critical for the inference of homology and phylogeny. Such explorations may be pursued in a variety of empirical research programs such as evolutionary developmental biology, comparative morphology, ontogenetic studies, and genetics. In addition, guides such as secondary structure, topology, connectivity, and function are all legitimate empirical ways to evaluate homology hypotheses. As in all empirical sciences, experiential and causal grounding of morphological character statements is unlikely to hit bedrock, but must rather be thought of as an explorative endeavor based on all available empirical avenues.

## Acknowledgments

We are grateful to an anonymous reviewer who provided helpful feedback. This work was supported in part by NSF grants DEB-0235618 and EF-0334961 to M.K and O.R.

## References

Audi, R., 2003. Epistemology. A Contemporary Introduction to the Theory of Knowledge, 2nd edn. Routledge, New York.

Ayer, A.J., 1952. Language, Truth, and Logic. Dover, New York.

Blanshard, B., 2001. Coherence as the nature of truth. In: Lynch, M.P. (Ed.), The Nature of Truth. Classic and Contemporary Perspectives. MIT Press, Cambridge, Massachusetts, pp. 103–121.

Carnap, R., 1950. Logical Foundations of Probability. University of Chicago Press, Chicago, Illinois. 607p.

Carnap, R., 1997a. Intellectual autobiography. In: Schilpp, P.A. (Ed.), The Philosophy of Rudolph Carnap. Open Court, La Salle, Illinois, pp. 1–84.

Carnap, R., 1997b. Replies and systematic expositions. In: Schilpp, P.A. (Ed.) The Philosophy of Rudolph Carnap. Open Court, La Salle, Illinois, pp. 859–1013.

Eernisse, D.J., Kluge, A.G., 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol. Biol. Evol. 10, 1170–1195.

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

Fitzhugh, K., 1997. The abduction of cladistics. Cladistics 13, 170–171.

Friedman, M., 1999. Reconsidering Logical Positivism. Cambridge University Press, Cambridge, pp. xix, 252.

Godfrey-Smith, P., 2003. Theory and Reality. An Introduction to the Philosophy of Science. University of Chicago Press, Chicago, Ilinois.

Grant, T., Kluge, A.G., 2004. Transformation series as an ideographic character concept. Cladistics 20, 32–31.

Griffiths, G.C.D., 1974. On the foundations of biological systematics. Acta Biotheor. 23, 85–131.

Haack, S., 1998. Evidence and Inquiry. Towards Reconstruction in Epistemology. Blackwell, Oxford.

Haack, S., 2000. A foundherentist theory of empirical justification? In: Sosa, E., Kim, E. (Eds.), Epistemology, an Anthology. Blackwell, Malden, Massachusetts, pp. 226–236.

Haack, S., 2003. Defending Science—Within Reason. Between Scientism and Cynicism. Prometheus Books, New York.

Hanson, N.R., 1961. Patterns of Discovery. An Inquiry into the Conceptual Foundations of Science. Cambridge University Press, Cambridge.

Härlin, M., 1999. The logical priority of the tree over characters and some of its consequences for taxonomy. Biol. Linn. Soc. 68, 497–503.
Hempel, C.G., 1965. Aspects of Scientific Explanation. Free Press, New York.
Hennig, W., 1950. Grundzüge einer Theorie der Phylogenetischen Systematik. Deutscher Zentralverlag, Berlin.
Hennig, W., 1966. Phylogenetic Systematics. University of Illinois Press, Urbana, Illinois.
Hennig, W., Schlee, D., 1978. Abriss der phylogenetischen Systematik. Stuttg. Beitr. Natkd. A 319, 1–11.
Hull, D.L., 1968. The operational imperative: sense and nonsense in operationalism. Syst. Zool. 17, 438–457.
Hull, D., 1994. Contemporary systematic philosophies. In: Sober, E. (Ed.), Conceptual Issues in Evolutionary Biology. MIT Press, Cambridge, Massachusetts, pp. 259–330.
Hung, T., 1992. Ayer and the Vienna Circle. In: Hahn, L.E. (Ed.) The Philosophy of A.J. Ayer. Open Court, La Salle, Illinois, pp. 279–300.
Kearney, M., 2003. Systematics of the Amphisbaenia (Lepidosauria: Squamata) based on morphological evidence from Recent and fossil forms. Herpetol. Monogr. 17, 1–74.
Kirkham, R.L., 2001. Theories of truth, a critical introduction. MIT Press, Cambridge, Massachusetts.
Kitcher, P., 1993. The Advancement of Science. Science Without Legend, Objectivity Without Illusions. Oxford University Press, Oxford.
Kluge, A.G., 1989. A concern for evidence, and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38, 7–25.
Kluge, A.G., 1997. Testability and the refutation and corroboration of cladistic hypotheses. Cladistics 13, 81–96.
Kluge, A.G., 1998. Sophisticated falsification and research cycles: consequences for differential character weighting in phylogenetic systematics. Zool. Script. 26, 349–360.
Kluge, A.G., 2001a. Philosophical conjectures and their refutation. Syst. Biol. 50, 322–330.
Kluge, A.G., 2001b. Parsimony with and without scientific justification. Cladistics 17, 199–210.
Kluge, A.G., 2003a. The repugnant and the mature in phylogenetic inference: atemporal similarity and historical identity. Cladistics 19, 356–368.
Kluge, A.G., 2003b. On the deduction of species relationships: a précis. Cladistics 19, 233–239.
Kluge, A.G., 2004. On total evidence: for the record. Cladistics 20, 205–207.
Lecointre, G., Deleporte, P., 2004. Total evidence requires exclusion of phylogenetically misleading data. Zool. Scripta 34, 101–117.
Leplin, J., 1997. A Novel Defense of Scientific Realism. Oxford University Press, Oxford.
Lipton, P., 2004. Inference to the Best Explanation, 2nd edn. Routledge, London.
Lynch, M.P., 2001. The Nature of Truth. Classic and Contemporary Perspectives. MIT Press, Cambridge, MA.
Naylor, G.J.P., Adams, D.C., 2001. Are the fossil data really at odds with the molecular data? Morphological evidence for Cetartiodactyla phylogeny reexamined. Syst. Biol. 50, 444–453.
Naylor, G.P.J., Adams, D.C., 2003. Total evidence versus relevant evidence: a response to O'Leary *et al.* (2003). Syst. Biol. 52, 864–865.
O'Leary, M.A., Geisler, J.H., 1999. The position of Cetacea within Mammalia: Phylogenetic analysis of morphological data from extinct and extant taxa. Syst. Biol. 48, 455–490.
O'Leary, M.A., Gatesy, J., Novacek, M.J., 2003. Are the dental data really at odds with the molecular data? Morphological evidence for whale phylogeny (re)reexamined. Syst. Biol. 52, 853–564.
Oberdan, T., 1993. Protocols, Truth, and Convention. Rodopi, Amsterdam.
Osborne, J.W., 1978. Morphogenetic gradients: fields versus clones. In: Butler, P.M., Joysey, K.A. (Eds.), Development, Function, and Evolution of Teeth. Open Court, La Salle, Illinois, pp. 171–201.
Patterson, C., 1982. Morphological characters and homology. In: Joysey, K.A., Friday, A.E. (Eds.), Problems of Phylogenetic Reconstruction. Academic Press, London, pp. 21–74.
Popper, K.R., 1973. Objective Knowledge, an Evolutionary Approach. Clarendon Press, Oxford.
Popper, K.R., 1979. Die beiden Grundprobleme der Erkenntnistheorie. J.C.B. Mohr (Paul Siebeck), Tübingen.
Popper, K.R., 1989. Conjectures and Refutations. Routledge & Kegan Paul, London.
Popper, K.R., 1992. The Logic of Scientific Discovery. Routledge & Kegan Paul, London.
Popper, K.R., 1997. The demarcation between science and metaphysics. In: Schilpp, P.A. (Ed.), The Philosophy of Rudolph Carnap. Open Court, La Salle Illinois, pp. 183–226.
Remane, A., 1952. Die Grundlagen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetic. Akademische Verlagsgesellschaft, Leipzig.
Richards, R., 2002. Kuhnian values and cladistic parsimony. Perspect. Sci. 10, 1–27.
Richards, R., 2003. Character individuation in phylogenetic inference. Philos. Sci. 70, 264–279.
Rieppel, O., 1992. Homology and logical fallacy. J. Evol. Biol. 5, 701–715.
Rieppel, O., 1994. Homology, topology, and typology: the history of modern debates. In: Hall, B.K. (Ed.), Homology: the Hierarchical Basis of Comparative Biology. Academic Press, London, pp. 63–100.
Rieppel, O., 2003a. Popper and systematics. Syst. Biol. 52, 259–271.
Rieppel, O., 2003b. Semaphoronts, cladograms, and the roots of total evidence. Biol. J. Linn. Soc. 80, 167–186.
Rieppel, O., 2004. The language of systematics, and the philosophy of "total evidence". Syst. Biodiversity, 2, 9–19.
Rieppel, O., Kearney, M., 2001. The origin of snakes: limits of a scientific debate. Biologist 48, 110–114.
Rieppel, O., Kearney, M., 2002. Similarity. Biol. J. Linn. Soc. 75, 59–82.
Ruse, M., 1988. Philosophy of Biology Today. State University of New York Press, Albany, New York.
Schlick, M., 1959a. The foundation of knowledge. In: Ayer, A.J. (Ed.), Logical Positivism. The Free Press, New York, pp. 209–227.
Schlick, M., 1959b. Positivism and realism. In: Ayer, A.J. (Ed.), Logical Positivism. The Free Press, New York, pp. 82–107.
Schultze, H.P., Trueb, L., 1991. Origins of the Higher Groups of Tetrapods. Controversy and Consensus. Comstock Publishing Associates, Ithaca, New York.
Scott, J.H., Symons, N.B.B., 1977. Introduction to Dental Anatomy. Churchill Livingstone, Edinburgh.
Siddall, M.E., Kluge, A.G., 1997. Probabilism and phylogenetic inference. Cladistics 13, 313–336.
Sneath, P.H.A., Sokal, R.R., 1973. Numerical Taxonomy. W.H. Freeman, San Francisco, California.
Sober, E., 1988. Reconstructing the Past. Parsimony, Evolution, and Inference. MIT Press, Cambridge, Massachusetts.
Sokal, R.R., Sneath, P.H.A., 1963. Principles of Numerical Taxonomy. Freeman, San Francisco, California.
Stroll, A., 2001. Twentieth Century Analytic Philosophy. Columbia University Press, New York.
Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–9.