

Empirical, cognitive, and truth-functional implications of considerations of parsimony in phylogeny reconstruction.

Richard Wikander, Department of Anthropology, University of Massachusetts at Amherst, Amherst, MA 01003.

Received 18 February 1989; 6 December 1990

ABSTRACT: The principle of parsimony has figured prominently in some approaches to phylogeny reconstruction as a justification for the acceptance or rejection of hypotheses of phylogeny. Maximally parsimonious phylogeny reconstructions are often regarded as more highly corroborated, by virtue of their simplicity, than less parsimonious alternatives. This is an indefensible use of parsimony, because it relies upon an interpretation of the principle which is itself indefensible. Maximally parsimonious reconstructions may in fact be less testable than their more complex competitors, and thus may have less information content. If the hypotheses of homoplasy which phylogeny reconstructions entail are genuinely untestable, then these reconstructions may themselves be untestable, and thus not meaningful scientific statements.

Introduction

The intent of this paper is to discuss some of the controversies surrounding the use of the principle of parsimony in phylogeny reconstruction. This is a familiar and on-going debate: some have argued that phylogeny reconstruction in accordance with the principle is neutral with respect to particular theories of the evolutionary process (e.g., Gaffney, 1979); makes no assumptions about nature (e.g., Farris, 1970, 1979, 1983); and yields results which are more "highly corroborated" or "natural" by virtue of their simplicity (e.g., Engelmann and Wiley, 1977; Kluge and Farris, 1969). The other side takes issue with at least some of these claims (see esp. Cartmill, 1981, on the unlikelihood of nature being "simple" in the postulated manner, and on the Wagner equivalence of alternative cladograms; Panchen, 1982, on a number of epistemological difficulties attendant upon using parsimony as a criterion of hypothesis selection; and Felsenstein, 1978, on possible worlds). I would like to address what Sober (1983) has called the "normative" problem of parsimony, or in other words the question of the appropriateness of appeals to parsimony to arbitrate between competing phylogenies with regard to their relative likelihood. Sober expressly restricted himself to the "descriptive" problem, or in other words how parsimony is actually used, noting with reference to the former problem only that there are serious unresolved problems attendant upon considering parsimony as a justification of theory acceptance or rejection.

Phylogenetic Systematics in Principle and Practice

I would, at this point, like to provide a brief description of the logical structure of phylogenetic systematics (hereafter referred to as PS), and of how it is used. I would also like to enter a note of caution here: PS is not so much a single method as a set of related methods; in other words, it is more properly considered as a methodology. Consequently, any attempt to provide a logical and procedural synopsis is bound to arouse some dissent from practitioners of some of its variants. I believe that the fundamental distinctions between types of traits and the procedure which I will outline are common to virtually all forms of PS. However, the axioms which I will suggest are not: these are restricted to those forms of PS which presume evolution as a part of their theoretical structure. This is in contradistinction to what is termed "pattern" cladistics, a form of analysis which groups synapomorphies rather than organisms per se. Pattern cladism does not, according to its adherents, presuppose evolutionary histories. Rather, it simply groups relationships between homologies

*

*

*

Evolutionary Theory 10: 33-46 (January, 1991)

The editors thank R. Johnson and A.L. Panchen for help in evaluating this paper.

© 1991, The University of Chicago

(cf. Panchen, 1982; Charig, 1982; for more in-depth discussions of pattern cladistics).

In preferring an evolutionary interpretation of PS, I follow Cracraft (1981, p. 460) when he notes that "... if classifications are to be used to investigate biological patterns in the fossil record, then those classifications must express the group membership of monophyletic groups, i.e., those assumed to have had a real genealogical history. Any classificatory procedure ... which creates groups that are not monophyletic will confound analysis of these biological patterns because some of the groups may not have a unitary genealogical history --- and it is the real history, after all, that we are attempting to understand."

Logical Structure of Phylogenetic Systematics

For the sake of brevity, I will assume familiarity on the part of the reader with the key terms of PS; full discussions of them may be found in any of a large number of publications (e.g., Joysey and Friday, 1982; Farris, 1983). I will emphasize once more, however, that I am concerned here with evolutionary PS.

All logical systems have an underlying set of axioms, and PS is no exception. I will not attempt to present an exhaustive list of them, because it is unnecessary to do so, and is probably impossible in practice if not in principle. However, I will offer several that I consider to be among the most fundamental. I will start with two that Gaffney (1979) has discussed, and add a few others.

Gaffney's axioms, in brief, are: 1. evolution has occurred; 2. new taxa are often characterized by new features. I would add at least the following four, perhaps as corollaries to the above: 1.1. there is genetic continuity within clades; 1.2. evolutionary change is correlated with genetic alteration; 1.3. morphology is largely under genetic control; 2.1. new features that characterize new taxa have one and only one genetic correlate.

As will become clear in the course of the present discussion, certain difficulties arise with regard to 1.3 and 2.1 above, though the need for these axioms should be equally clear. However, granting the above, only some subset of the total characters of the morphotype of a taxon can be used to uniquely characterize that taxon, and this subset is comprised of apomorphous traits. If there cannot be found at least one trait of this type, then the taxon is invalid (paraphyletic or polyphyletic). If at least one such trait can be found, then the taxon is monophyletic. Tracing the distribution of such traits among organisms allows the resolution of these organisms into groups, and allows the relative placement of these groups into larger ones, to yield nested hierarchies. The basic component of these hierarchies, however, is the "3-taxon" (or, more properly, "3-taxa") statement. Thus, three low-level taxa A, B, and C may, depending upon the distribution of traits, be resolved into ((AB)C), ((AC)B), or ((BC)A). Such statements are usually represented diagrammatically by cladograms, which are essentially evolutionary "trees" from which the temporal axis has been removed, so that evolutionary, but not temporal, distance is charted.

The logical structure of a science may be analyzed as either an uninterpreted axiomatic system, an interpreted system, or may be operationalized for application to practical problems. For instance, geometry may be treated as a purely abstract system of algebraic statement forms embodying relations only, with its terms unspecified or uninterpreted. It may also be expressed diagrammatically, as in analytic geometry, and in this case is an interpreted system, since its terms have some reference, however abstract. It may also be coupled with some set of correspondance rules which allow its application to the "real" world, rules which prescribe understanding "straight line," for instance, as "ray of light" or "taut rope" or some such; or "point" as "sharp end of a pin" or "dot on a printed page." The correspondance rules are what render theoretical science empirical and applicable to practical problems (see Nagel, 1961, pp. 131-137. It is not generally agreed that a theory need be viewed as an axiomatic system; my point is that it can be, and this view can be analytically useful).

Perusal of the pages of journals such as *Systematic Zoology* or *Cladistics* will reveal page after page of abstract statements in the form of computer-generated and solved algorithms for the resolution of 3-taxon statements. This is PS as an uninterpreted system. Accompanied by diagrams and explicit references to organisms and morphological or behavioral features, PS becomes an interpreted system, especially if those references are to hypothetical and somewhat idealized examples. Applied to real systematic problems, PS becomes of course a system for solving problems in natural science. In order to achieve such operationalization, rules of correspondance are required.

Practical Problems

As frequently happens when theoretical systems are applied to empirical matters, certain difficulties crop up in the attempt to use PS in the reconstruction of phylogenies and the resolution of taxonomic relationships. Among those problems which I consider to be the most serious is that of distinguishing homology from homoplasy. Features which conform to axioms 2 and 2.1 above are synapomorphies, and they obviously must be distinguished from homoplasies. The ability to make this distinction is a sine qua non of PS, but devising a reliable means of doing so has proven to be a surprisingly recalcitrant problem. The reason for this is that homoplasy is not simply what I will here refer to as "naive analogy," the "insect wing/bat wing" kind of comparison with which we are all familiar. "Homoplasy" includes traits which can be for all intents and purposes virtually identical in morphology and function. In theory, there can be only one "correct" resolution of relative closeness for any three taxa, simply because a given group of organisms cannot have more than one evolutionary history. And since a distinct group of organisms must be distinct according to something, and since that "something" can be neither symplesiomorphous nor homoplasious, it necessarily follows that any given evolutionarily "natural" group must be characterized by at least one synapomorphy. Thus, if one hypothesizes the grouping ((AB)C), and either A and C or B and C prove to share a synapomorphy within this 3-taxa grouping, then the hypothesized grouping has been falsified. Unfortunately, this seems to happen with disturbing frequency for any grouping of taxa A, B, and C; in other words, there is very often at least one incongruent character, so that all possible groupings are falsified. However, this latter situation is impossible within the framework of current evolutionary theory. One of the possible groupings must be correct. This dilemma is not, so far as I can tell, a shortcoming in the logical structure of PS; rather, it appears to be an unfortunate and somewhat unexpected practical limitation which seems traceable to deficiencies in background knowledge. At this point, a few examples are in order. I will take the first two examples from my own research, and draw several others from the literature.

The marmoset Callithrix jacchus and the common eastern grey squirrel Sciurus carolinensis share a strikingly similar distal tibiofibular articulation, and C. jacchus is quite different from tamarins in this regard (Wikander, 1987). However, because of the enormous evolutionary distance separating marmosets and squirrels, we unhesitatingly take this similarity to be homoplasious rather than homologous. However, Callithrix also shares this joint configuration with Saimiri sciureus. In the latter situation, one feels a bit more hesitant about drawing conclusions regarding the nature of the similarity. The reason, of course, is that marmosets and squirrel monkeys are considerably closer in relationship than either is to squirrels. But does evolutionary closeness render a similarity more likely to be homoplasious, because closely related animals are more likely to independently come up with the same solution to a given adaptive problem, or homologous, because of proximity to common ancestry? The answer is that there is no single, simple answer to the question. In this particular case, we are likely to retain Callithrix within the Callitrichidae, because of the weight of its numerous resemblances to other callitrichids. However, such a procedure, although intuitively attractive, comes perilously close to numerical taxonomy, a procedure which most cladists reject. In PS, what is important is not sheer number of similarities, because this does not take into account symplesiomorphies and homoplasies; rather, PS considers only resemblances in "good" traits, in this case "good" meaning apomorphous.

One good trait must, in the logic of PS, outweigh a hundred bad ones. In order to do this, it is essential that a verdict of "good" or "bad" be reached beforehand, so that it guides the construction of hypotheses of relationship, and that the verdict not be reached on the basis of some previously accepted hypothesis of relationship.

As a second example, a supernumerary bone referred to as the "prehallux," essentially a sesamoid occurring in the hallucial metatarsophalangeal joint in several orders of tetrapods, has been used to at least preclude close evolutionary relationships between some primates (see Lewis, 1964, 1972). However, it has been argued (see Wikander et al., 1984, 1986) that the genetic factors governing the appearances of sesamoid bones have yet to be separated from the purely biomechanical influences, and that the functional significance of the prehallux cannot be the same in all animals which possess it. Thus, the soundness of phylogenies constructed or precluded on the strength of this feature cannot be assessed until the appropriateness of the feature has been.

The traits in the above examples may seem poor ones because they are such simple ones, features which could easily appear independently in all sorts of vertebrates regardless of relationship, especially in the presence of selective pressures. It is possible that relatively complex characters could provide more reliable markers of relationship, since the probability of parallel evolution of complex states may be low. Unfortunately, virtually nothing is known of the relative probabilities governing the homoplasious appearances of complex and simple traits, and as a result, complexity does not carry much weight in practice, or at the least, that weight is subject to interpretation. For instance, Pettigrew (1986) discovered a remarkable similarity between primates and a genus of megachiropteran in the visual pathway to the brain. Since the condition was previously thought to be synapomorphous within primates, the clear implication of this discovery was that some megachiropterans are actually primates. However, R. D. Martin (1986) argued that the weight of resemblances between micro- and megachiropterans is such that Pettigrew had uncovered a stunning homoplasy, regardless of its complexity. On the other hand, in a review of Schwartz (1988) on cladistic relationships within the Hominoidea, L. Martin (1988, p. 58; no relation to R. D. Martin) argued that the claim that chimpanzees and gorillas are more closely related to one another than either is to Homo is supported not only by molecular evidence but also by "compelling morphological evidence ... the two species share all the specialized features of the forelimb associated with knuckle walking: the bones of the fingers and palm, as well as those in the wrist, are tightly packed and bound with ligaments, so that the wrist and hand can operate together as a stable unit; the lower end of the radius ... is thickened for extra strength and grooved in such a way as to allow the wrist to lock into position; and the front surfaces of the fingers, which land on the ground in knuckle walking, are reinforced with thick skin pads. Such close similarity in a complex set of structures seems unlikely to have evolved independently in each species" (italics added).

I believe that Callithrix is a callitrichid, but Saimiri is not; I also believe that both Callithrix and Saimiri are primates, but that Sciurus is a rodent. I also suspect (although I am not certain) that no bats are really primates. I have no idea as to the proper resolution of the tripartite split between Pan, Homo, and Gorilla. The point which I would like to emphasize here is that "goodness" and "badness" of traits is in practice (though not in theory) rather intuitive, and a contextual matter, determined frequently on the basis of an overall gestalt. There are no rules of correspondance which prescribe the unambiguous translation of the theoretical terms of PS into empirical ones. This is especially true for the terms "homology" and "homoplasy." We all know what the phrase "homologous character" means: we understand its intension. However, we cannot find such a character with real reliability in many instances: in other words, we encounter considerable difficulty in specifying its extension. PS has reached the stage of an interpreted axiomatic system, but its rules of correspondance are still incompletely formulated. The principle of parsimony has been made to bear this burden in the place of such rules.

Realism vs. Non-Realism in Science

My interest in the topic of parsimony in phylogeny reconstruction stems from my interest in the broader debate, within the philosophy of science, between realists and non-realists, viz., between those who claim that science somehow captures or more or less accurately encodes some aspect of a reality which is independent of the minds which seek to know it, and those who deny this thesis. The connection between this debate, and my present topic, may become clearer by consideration of the following.

The debates over the broad issue of scientific realism, and over the narrower issue of the use of parsimony in phylogeny reconstruction, appear to involve some of the same basic epistemological considerations: 1. empirical content; 2. cognitive claims; and 3. truth content. Briefly, and in order, these are matters of, first, empirical consequences, since every "good" scientific theory must, even if indirectly, entail some observable consequences; second, knowledge claims, since most (though probably not all) scientific theories are interpreted as implying or conveying some kind of knowledge about the "real" world; and third, justifying these knowledge claims, in the sense that a realist, at least, invokes some notion of "approximate" truth in order to justify adherence to particular theories, and to account for the efficacy of those theories which do, at least provisionally, appear to work. Realists and non-realists disagree mainly over the second and third of these considerations, the latter claiming that empirical content cannot arbitrate claims of knowledge and truth, since for a given set of empirical data, there is an indefinitely large number of theories, the internal structures of which are inconsistent with one another, which will satisfactorily account for those data (see Leplin, 1984; and Cushing, Delaney, and Gutting, 1984, for a large number of papers presenting a broad spectrum of variants of the realist and non-realist positions). This is the "evidential indistinguishability" thesis, and we will have occasion to return to it. Truth claims are out of place here, since mutually inconsistent and even contradictory theories cannot all be "true," and as a result, knowledge claims can neither be supported nor justified.

Phylogeny reconstruction must deal with these same issues: what do theories of phylogeny immediately account for, and on what basis? This is a question of empirical content. Also, how do we regard these things we call "reconstructions": do we think of them as allowing us to "know" anything, in the sense of rationally justified belief, or are they simply more or less efficacious tools or devices for ordering things, just as hammers or screwdrivers are more or less efficacious tools for performing certain tasks? And finally, can we determine if these "reconstructions" bear any sort of resemblance to the "real" history of the entities involved, or is "truth content" a meaningless or confusing notion to introduce into the field? Different approaches to systematics begin to sound, in this context, somewhat like different approaches to the question of scientific realism, with phylogeneticists perhaps on the realist side, and pattern cladists on the other. I do not wish to press this analogy, and I do not mean to imply that "pattern cladism" is a form of non-realism; I do, however, believe that there are differences in at least some of the cognitive claims made in different approaches to systematics, and that these differences may reflect some divergence of attitudes toward ontology and truth content.

I suspect (though I am not certain) that most systematists, like most scientists in general, are realists, in that they consider themselves to be engaged in the attempt to try to capture some piece of a reality which exists independently of them, and it is this approach with which I am concerned. With regard to the use of parsimony in phylogeny reconstruction, truth content and cognitive claims, or in other words the extent to which one is justified in placing confidence in whatever parsimony-grounded knowledge claims reconstructions may be taken as implying, are the central considerations. Claims of high corroboration or naturalness seem to imply a truth-functional role for the principle of parsimony. Is this position defensible?

Parsimony in the Philosophy of Science

One's definition of parsimony will depend upon one's field and purposes. In systematics, the prevailing definition utilizes the notion of minimizing ad hoc hypotheses of homoplasy (Farris, 1983). Engelmann and Wiley (1977) refer to these hypotheses as "statements of error," and consider them to be untestable (see also Farris, 1983). This approach to homoplasy is apparently still characteristic of PS, at least implicitly, and it is this general, and still current, approach which is the focus of the present study.

When a scientist attempts to render intelligible a certain phenomenon or range of phenomena, one of two situations presents itself: rarely, the scientist has no idea at all of why or how things are the way they are; or usually, the scientist has too many ideas. The first problem is one of creativity, the second of decision, of narrowing the field of potential explanations to a manageable few. In either case, one way to approach the problem would be to devise some reasonably objective (though perhaps arbitrary) criteria for solving it.

Because of the enormous difficulties attendant upon deriving prescriptive guidelines for the initial construction of hypotheses (the "logic of discovery" problem; see Hanson, 1958; Salmon, 1966; Laudan, 1981), much of the philosophy of science has concerned itself with what might loosely be termed problems in the logic of justification, such as inferential structure, the nature of test events, the ontological status of theoretical entities and relations, various extralogical considerations, and so on. Among these problems is the decision problem mentioned above: narrowing the field of potential explanations. This is not the same as the problem of eliminating some one or more from a group of competing hypotheses via testing; rather, it is prior to this.

Both of these problems are concerned with questions of acceptability, but in two different senses. If a researcher finds it necessary to eliminate some hypotheses from consideration due to a relative surfeit of them, at least some of the following (non-exhaustive) factors are likely to influence the results of this decision.

First, there is the matter of time and resource management. Considerations of time, money, equipment, and competing responsibilities will virtually always place relatively severe constraints on a researcher's ability and inclination to pursue problems. Second, there are plausibility considerations. It is extremely unlikely that all hypotheses available for consideration will be of equivalent plausibility. Third, there is the question of interest. Not all available explanations will be of equal interest to the investigator, or to the scientific community to which he or she belongs. Fourth, there is the matter of relevance. There are many different ways in which to ask and answer questions. For example, morphological features of organisms may be explained on the basis of adaptive advantage, with appeals to selective pressures, or on the basis of what might be termed internal causal mechanisms such as allometry, heterochrony, or whatever. Any of these approaches may be of interest, but not all may be germane to the specific problem at hand.

The above considerations are obviously intertwined. For instance, consideration of time and resource management is likely to be applied on a case-by-case basis, but no amounts of time and money are likely to inspire a researcher to pursue patently absurd or utterly uninteresting ideas, or to pursue problems which are either intuitively or demonstrably insoluble. The decision regarding which hypotheses are worth the trouble to examine will be based on a complex mixture of these factors, and many others as well. Among some of the other factors which may, depending upon circumstance, influence the decision, are ease of testing, generality, precision, analogy (i.e., similarity to other, successful hypotheses), and simplicity of parsimoniousness. None of these factors are test conditions; hypotheses which they eliminate have not had any substantive, factual claims evaluated or falsified.

Acceptability in a different sense arises within the context of testing. Once an investigator has settled upon a set of hypotheses to be examined via testing, the subsequent acceptability of some one or more of these hypotheses rests on the successful passing of such tests. And unless the investigator is thoroughly non-realist

in outlook, and regards assertions of truth content as meaningless or superfluous, the testing procedure will carry with it an implicit assertion of such truth content. At the very least, the investigator will have in mind some notion of verisimilitude, whereby surviving hypotheses are considered to at least approximately encode relations and entities which in fact obtain in the ontologically independent, "real" world, whereas unsuccessful hypotheses do not.

Acceptability in the sense of determining the initial selection of hypotheses to be examined carries with it no such implicit assertion of "truth," even in an approximate sense. An investigator may reject a hypothesis because it too suspiciously resembles one which has failed in the past, but this suspicion, while perhaps well-founded from the scientist's point of view, is obviously in and of itself not an evaluation of the empirical claims made by the hypothesis. Depending upon the precise nature of the resemblance, the past failure of the one need not in any way reflect upon the potential for success of the other; the fates of the hypotheses may well be entirely independent. If truthlikeness is entertained at all for a hypothesis, it is entertained only within the context of the test procedure, since test results, but not psychological states (e.g., suspicion), are intersubjectively shareable. Thus, if one says of a hypothesis that it is "not likely to be true," based on analogy with similar hypotheses which have failed, this is a different kind of truth assertion from the rejection as probably false of a hypothesis as a result of the outcomes of tests. "Similarity" in the sense being discussed here is simply too subjective a concept to be relevant to the problem of the rational justification of hypotheses.

Parsimoniousness or simplicity has been regarded by various philosophers as a desirable feature for hypotheses to exhibit, and one which will, in some situations, render some hypotheses preferable to others. In order to increase the corroboration of any hypothesis, however, and specifically phylogeny reconstructions, it must lend to that hypothesis an acceptability in the second sense discussed above, namely as a test condition which implies assertions of truth content. In order to do this, it must carry with it implications for ontology.

The principle of parsimony may be understood in two distinct, but not necessarily mutually exclusive, senses. I have characterized these elsewhere (Wikander, 1985) as the "ontological" and the "heuristic" senses. Ontological characterizations may be taken to be those formulations of the principle which refer to or may be taken to imply assertions about the world or nature, such as "nature does not multiply entities unnecessarily," or "nature proceeds to a given result in the simplest manner possible." Heuristic formulations are those which are prescriptive in character, such as "one (i.e., the investigator) ought not to multiply entities unnecessarily," or "the simplest among competing solutions to a given problem is to be preferred."

The differences between these two types of statement are essentially matters of ontological import, cognitive content, and truth assertions. Ontological formulations imply statements about ontology, and make claims about world structure and the relations which obtain between that world and the conceptual structures which are intended to describe it. They make knowledge claims, and as such, they imply truth content. Heuristic formulations, as noted above, are prescriptive, and are justified most immediately on aesthetic or practical grounds. They need not be taken as implying any particular ontology, they make no cognitive claims, and they have no truth content.

Generally speaking, the philosophy of science has understood the principle in the heuristic sense, and thus as a canon of acceptability in the first sense discussed above. It is an environmental consideration, in that it is dependent upon the theoretical environment or context. Given two or more competing hypotheses which do not differ in any relevant evidential manner, but which differ in their relative degrees of complexity in terms of internal structure and articulation with the environment, the principle sanctions the choice of the simplest, all else being equal. Deciding what is to be minimized or simplified in a theory is itself an environmental consideration, as Sober (1983) points out with reference to systematics. As noted above, it can be in terms of inter- or intratheoretical articulations; it can also be in terms of the

nature of the internal theoretical structure, with regard to the entities postulated, the relations between these entities, or both; or, and very frequently (perhaps usually), it can refer to the degree of synthesis between previously available theories which the theory in question is able to effect.

It is commonplace for scientists to express the belief that nature is indeed simple. But "simplicity" as applied to any particular theory is always an a posteriori assessment, a property of the theory relative to its efficacy in doing what is intended of it within its environment. To argue that relativistic physics, for example, is in any absolute sense "simpler" than classical mechanics is plainly false, but its synthetic power is in many instances considerably greater. The attribution of parsimoniousness to theories is the result of hindsight, and somewhat anachronistic. And in any event, Laudan (1984, p. 223) has argued that, depending upon one's criteria, the history of science may be seen to be marked more by repeated failure than by stunning success; it would follow from this that parsimoniousness has, at least historically, been a suspect guide to verisimilitude, even in retrospect. Let us suppose, however, that historical records are too uncertain to be reliable in the matter. Let us further suppose that the principle of parsimony can be understood in its ontological sense. What would it assert, and how might the assertion be established?

The Ontological Interpretation and Potential Justifications

Very briefly, the principle of parsimony would assert that of all P possible worlds, there is a certain subset N of these, where $1 < N < P$, such that N is comprised of those possible worlds which are simpler than the rest according to some standard. Ideally, we would like $N = 1$, but there is no reason for believing that this need be so; there may be more than one world of comparable simplicity. We can, for the sake of argument, understand "simpler" in any manner we choose.

There are a number of different ways in which we might attempt to establish such a statement; we will examine in a bit of detail how one might make the attempt via what I will call a "transcendental" argument, and then briefly discuss some alternative approaches. In doing the latter, we will follow Watkins' (1984, pp. 93-104) excellent discussion of the utility of these forms of argumentation in attempting to establish the validity of some form of inductive principle (see also his own interesting discussion of attempts to establish some form of parsimoniousness in order to justify an inductive principle, pp. 89-116 of the above-cited work).

We can begin with some unspecified theory T . We will assume that T actually "works," in the sense that, for some input I consisting of various initial conditions, and some output O which is at least in principle observable, and much of which does appear to be empirically corroborated, the following relation holds: $T(I) = O$.

This relation does not uniquely determine T : there are in fact an indefinitely large number of T s which will also satisfy the relation, in accordance with the evidential indistinguishability thesis. With regard to O , then, there is a set L , $L = [T_1, T_2, \dots, T_n]$ consisting of evidentially indistinguishable elements satisfying the relation.

Each T_i is ex hypothesi distinct, and since this distinction cannot be on evidential grounds, it must be in terms of some internal structure S_i . This structure can be understood to be the theoretical entities which are referred to in T_i , the relations which are specified to obtain between these entities, or some other relevant characteristic of scientific theories.

Let us imagine that L contains some one simplest element, and designate it by T_s , and its structure by S_s . For any T_i and S_i , where $i \neq s$ and "<" is taken to mean "simpler than," the following relation holds: $T_s(S_s) < T_i(S_i)$. Does T_s have any relevance for our particular world, in the sense that S_s more accurately captures its structure than does any other S_i ?

We already know that T_s , like the others, is evidentially satisfactory, so that the required comparison between our world and T_s must be carried out via an examination of the internal structures S_s and S_w , where S_w denotes the actual structure

of our world. However, the so-called "God's eye view" problem prevents us from directly intuiting S_w , and is in fact what causes our current problem in the first place. The only approach to the problem which does not rely on mysticism is to formulate some hypothesis H which we presume captures S_w . However, we cannot use evidential means to construct one unique H , because of the evidential indistinguishability thesis; rather, we will have a set $M = [H_1, H_2, \dots, H_n]$ which constitutes our conjectures about S_w . Since L consists of all possible worlds satisfying the evidential above, M will be a subset of L , and we will find ourselves comparing the elements of L with one another. We will find some H_s , but $H_s = T_s$, and we will be led back to our original problem. This situation will obtain for any specific form of simplicity we choose to examine; the degree of simplicity manifested by our world cannot be determined without complete and certain knowledge of S_w , in which case the problem with which we began does not arise.

As noted above, there are other types of argument whereby we might be able to justify the ontological interpretation of the principle. I will briefly discuss them here, explaining why I think they cannot work.

1. Empirical justification. According to this view, the simpler among competing hypotheses tend, as a matter of historical fact, to work as well as or better than their more complex competitors. In fact, parsimony can be either denied any empirical confirmation or rescued from any disproof via its local character. For any falsification, one can always appeal to over-all or global parsimoniousness by arguing that inclusion of a broader sample of reality will demonstrate such over-all simplicity. Any corroboration can be invalidated via the same means. And in any event, the preceding discussion should make it clear that appeals to evidence cannot serve to settle the dispute.
2. Vindication. This form of argumentation has been used principally by those concerned to justify some sort of inductive principle. It seeks to establish that if any method works or is valid, then the particular one in question will work best. Sober (1983) very briefly touches on this with respect to parsimony in systematics. Such an approach begs the question, it seems to me, because the issue is to establish that in fact something works or is the case, and not simply that it would work, or hold true, if certain conditions were met. And as Watkins (1984, pp. 99-100) points out, Carnap has shown that however superior one inductive principle might be in comparison with others, there might well be possible worlds in which the others perform better. And this is exactly what Felsenstein (1978) has shown with reference to parsimony in phylogeny reconstruction.
3. Self-justification. Parsimony justifies itself by virtue of the fact that it works. Two points may be made here: first, this is precisely the issue in question; second, an appeal to past success collapses into an appeal to empirical justification, which has already been discussed.

The only option remaining is that the principle of parsimony is a synthetic a priori truth, or in other words is both a statement about the world and is necessarily true. However, this is not really an argument, but rather a claim that the point at issue is impossible to deny, in the same way that axioms of Euclidean geometry were thought to be impossible to deny of the world of experience. Parsimony within the context of phylogeny reconstruction is clearly not such a statement, by virtue of the debate which surrounds it. And the fortunes of Euclidean geometry have demonstrated the wisdom of caution in claiming necessary truth for even the most apparently "self-evident truths."

As noted earlier, simplicity is a contextual attribute. All theories are local at some level, the only one not so being the still unattained, and possibly unattainable, theory encompassing all phenomena. For any local theory T_s , there will be some more "global" simple theory GT_s which accounts for the same phenomena, and more besides. Often, it will prove impossible to fit T_s into GT_s without sacrificing the simplicity of the former. Under certain conditions, this may occur in phylogeny reconstruction, so that parsimony is not additive (Engelmann and Wiley, 1977; Maddison, Donoghue, and Maddison, 1984). If parsimony is taken as an indicator of verisimilitude, then the situation may arise in which it simultaneously asserts likely truth of a theory T_s

(including a phylogeny reconstruction), and likely truth of some theory GT_s . In the event that these theories do not articulate, at least one of them must be false, which is contradictory.

What, then, of the claim that the most weakly or infrequently falsified hypotheses are ipso facto the most highly corroborated? This appears to imply some sort of empirical evaluation of the theory itself, and sounds intuitively plausible and a potential justification of parsimoniousness. However, it contains a potentially serious and perhaps fatal flaw, as may be seen if we examine it from the perspective of a famous paradox in the philosophy of science, the "paradox of confirmation."

The Paradox of Confirmation

We will begin with a general view of the paradox, and consider an extreme example in order to illustrate its major features. What follows is a standard discussion of it, and treatments essentially the same but in more detail may be found in most logic or philosophy of science texts (e.g., Ackermann, 1976; Lambert and Brittan, 1970).

Consider the statement "All ravens are black," which we can write as $(Ux)(Rx \rightarrow Bx)$, where "R" means "raven," "B" means "black," and standard logical connectives and quantifiers are used. This is logically equivalent to the statement "There are no non-black ravens," or $\neg(Ex)(Rx \& \neg Bx)$.

Imagine that the universe can be divided into four quadrants, such that every entity in the universe can be unambiguously assigned to one and only one quadrant, as in the following:

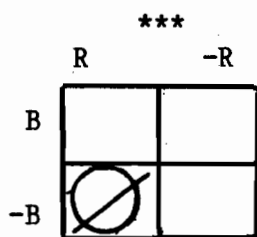


Fig. 1. All ravens are black.

According to this partitioning, any entity we choose to consider will be either a black raven (top left), a black non-raven (top right), a non-black non-raven (bottom right), or a non-black raven (bottom left). The bottom left quadrant, that occupied by non-black ravens, is empty according to our statement, so that every entity we observe must properly fall into one of the other three. One observation of an entity properly belonging in the bottom left quadrant is logically sufficient to falsify the hypothesis "All ravens are black." However, it is not true that every entity which fails to falsify it thereby corroborates it, for it follows from this position that white handkerchiefs, electromagnetic fields, pumpkin pies, and today's weather conditions on Mars will all count as corroborating evidence. In other words, every entity in the universe, with no restriction on how one defines the term "entity," which is not a non-black raven may figure in a "corroborative" observation. The paradox arises from the interpretation of every non-unfavorable (i.e., non-falsifying) observation as some sort of positive evidence in favor of the hypothesis in question. This is, however, how character distributions in systematics seem on occasion to be interpreted. For example, Engelmann and Wiley (1977, p. 5) state, with reference to the allegedly greater empirical content of hypotheses of ancestor-descendant relationship, that "... it appears that it is possible to refute an ancestor-descendant relationship and that every character of [taxon] B that is the same as [taxon] C or plesiomorphic relative to C fails to refute such a hypothesis and thus corroborates it" (italics added). They also remark that when all possible reconstructions have apparently been falsified via character distributions, the criterion of parsimony may be used "to select the hypothesis which includes the fewest internal inconsistencies ... Thus, the most parsimonious of two or more hypotheses is the most highly corroborated one" (p. 6; italics added).

The paradox is a consequence of the structure of the logic of confirmation, and in spite of their appeals to Karl Popper's philosophy of science, Engelmann and Wiley are in practice employing the logic of confirmation. One obvious source of the paradox is the unrestricted universe of discourse: it is intuitively ridiculous to consider the brand of one's toothpaste, for example, a confirming instance of "All ravens are black." But in fact it is, because the restriction of the universe of discourse must be accomplished by extralogical means. The paradox remains even if we restrict ourselves to, say, all birds and only birds, or even to birds of a certain size only. There is no logical reason why restricting observations to birds is any less arbitrary than restricting them to endotherms, or vertebrates, or whatever, unless this restriction is sanctioned by a set of auxiliary extralogical assumptions. Constructing such a set of assumptions in a relatively objective, non-intuitive way is, in empirical science, an extremely difficult if not impossible task.

The intent of the paradox of confirmation is to show the simultaneous weakness in practice of the logic of confirmation and the strength in principle of the logic of falsification: a limited universe of discourse is guaranteed by the falsifiers of a hypothesis, and the paradox makes it clear that the strength of a hypothesis is more a function of the number and types of its falsifiers than it is the weight of its "corroborating" instances. A hypothesis with no corroborating instances is, of course, of little use, but it is from its falsifiers that a hypothesis acquires its information content, and its strength derives from explicit attempts to assess the claims of these falsifiers: one cannot minimize their number and then disregard them. The principle strength of a hypothesis most emphatically does not derive from counting its corroborators.

This brings us to the significance of the "untestable" ad hoc hypotheses of homoplasy. Insofar as considering maximally parsimonious phylogeny reconstructions to be ipso facto more highly corroborated amounts to counting up confirming instances, this approach risks neglecting the full import of the putative falsifiers. Ad hoc hypotheses of homoplasy are the predictions of hypotheses of phylogeny, making explicit empirical claims of great importance for these hypotheses.

Ad hoc hypotheses, if they are indeed untestable, cannot serve as falsifiers. Falsifiers are themselves low-level hypotheses which must make empirical claims and have truth content; it is not enough that they be synthetic statements. If they are not testable, they have no discernible truth content (regardless of whether or not they "sound" plausible), their own falsifiers or test conditions are unascertainable, and as a result, their empirical claims cannot be assessed. There is little epistemological difference between a hypothesis rendered unfalsifiable via untestable ad hoc adjustments, and a hypothesis constructed so as to be untestable to begin with. If phylogenies are potentially falsifiable by incongruent character distributions, but these distributions are accounted for or explained away by genuinely untestable hypotheses of homoplasy, then the entire theoretical structure is rendered untestable. Justifying the ontological interpretation of the principle of parsimony might offer a means of salvaging this situation, is a sufficiently strong and specific formulation of it could be established. However, it appears to be impossible in principle to establish even a generous version of it. Granting this, if hypotheses of homoplasy are truly untestable, then they have no empirical content, and not only does the evidential indistinguishability thesis hold for hypotheses of phylogeny, but what is worse, these hypotheses are probably not even scientifically meaningful statements.

If the concept of homoplasy has empirical content, there ought to be ways in which it can be construed so as to have observable consequences, and which will in turn render testable those hypotheses in which it occurs. However, this does not guarantee that such testing would be easy to carry out. Homology and homoplasy must be distinguished on some basis other than gross morphological resemblance, and this distinction must, within the context of current evolutionary theory, be based on some statement of genetic identity itself based of necessity on phenetic criteria. However, except in very rare instances, such genetic identity cannot now be directly determined, and as a result, the statement of identity needs to be employed as a

sanctioning assumption covering some combination of methods of indirect investigation. Among such methods already in use are such things as the concordance of traits in the reasonably presumed absence of selective pressures which would make homoplasies likely; patterns of ontogenetic development; biogeographic and biostratigraphic data; and so on. There are obviously fairly serious methodological and theoretical problems still remaining to be dealt with in these areas, the central one perhaps being the still insufficiently understood connection between the genome and its phenetic expression. However, perfectly sound applications of this strategy exist (e.g., Novacek, 1980; Hartman 1986; Ward, 1986). One consequence of this discussion is the observation that, while it may be possible to construct phylogenies without reference to specifics of the evolutionary process (a point which many cladists have argued), it hardly seems possible to test these hypotheses in such a theory-neutral environment (cf. Brooks and Wiley, 1985).

Another point of interest to emerge from this discussion is the likelihood that, as a result of varying degrees of difficulty in assessing homoplasy in different kinds of features, relatively unparsimonious reconstructions may on occasion prove to be preferable if the hypotheses of homoplasy which they entail include characters which are more easily examined. A phylogeny requiring only one such hypothesis is clearly worthless if that hypothesis proves to be genuinely untestable: the phylogeny in effect has no empirical falsifiers. A relatively unparsimonious phylogeny implies many potential falsifiers, forbids more in this sense, and as a result, has a greater information content. Consequently, a reasonably sound case may be made for preferring unparsimonious phylogeny reconstructions in some situations.

Any procedure which we may choose for testing phylogenies, regardless of its specific nature, will invoke simplicity or parsimony in some sense. However, utilizing multiple independent methods for assessing homoplasy will utilize the pairwise neutrality if theories, in that there is no a priori reason why the results of such methods should agree. If we choose on the basis of such agreement we are, of course, choosing parsimoniously once again. However, it is parsimoniousness within a theoretical context which sanctions this interpretation of such agreement; the problem of justifying this use of parsimony is one which we do not face alone. It is a dilemma for all empirical science which lays claim to truth content of some sort, and takes a realist perspective in its approach to its subject, as a metascientific, or, if you will, a "metaphysical" problem. But I do not think that shifting the problem to this level is entirely a matter of begging the question. There are a great many other metascientific epistemological problems of just this sort, which are unsolved and perhaps insoluble; this is why there are so many non-realists. But if practitioners of particular sciences do not interpret such problems as precluding any sort of cognitive and truth functional claims for what they do, then it is both necessary and desirable that they at least ensure that intra-scientific methods and epistemology are consistent and sound.

Conclusion

In conclusion, I would like to reiterate three points which I consider to be among the most important:

1. Parsimony cannot be interpreted as a corroborative device, but rather only as a prior-to-testing decision criterion. To use it as a test condition amounts to leaving phylogeny reconstructions as plausible but completely unassessed hypotheses with undetermined relative likelihoods.
2. There can be no theory-neutrality in the testing of phylogeny reconstructions, even if the reconstruction itself can be done in a relatively theory-neutral manner.
3. Less parsimonious reconstructions may, at least in some situations, be preferable. If hypotheses of homoplasy are potential falsifiers, then maximally parsimonious reconstructions are in effect those which are maximally defended from falsification, which seems contrary to good scientific practice. And if these hypotheses of homoplasy cannot themselves be tested, then neither they nor the phylogenies in which they occur can be regarded as having information content, or scientific status.

Acknowledgements

I would like to thank Dr. A. L. Panchen and a reviewer who did not expressly waive anonymity for their extremely valuable suggestions for improving an earlier version of the present work, and H. H. Covert and A. J. Petto for a great many stimulating conversations which in various ways had a positive influence on this paper. I would also like to thank D. H. Chitty, Emeritus Professor of Zoology at the University of British Columbia, Vancouver, for initially suggesting the terms "ontological" and "heuristic" to refer to interpretations of the principle of parsimony. This acknowledgement was inadvertently omitted from an earlier work, and I would like to take this opportunity to rectify the oversight.

References

- Ackermann, R. J. 1976. Karl Popper. University of Massachusetts Press, Amherst.
- Brooks, D. R. and E. O. Wiley. 1985. Theories and methods in different approaches to phylogenetic systematics. *Cladistics* 1:1-11.
- Cartmill, M. 1981. Hypothesis testing and phylogenetic reconstruction. *Zeitschr. Zool. Syst. Evolutionsforsch.* 19:73-96.
- Charig, A. 1982. Systematics in biology: a fundamental comparison of some major schools of thought. In K. A. Joysey and A. E. Friday (eds.), Problems of Phylogenetic Reconstruction. Academic Press, New York, pp. 363-440.
- Cracraft, J. 1981. Pattern and process in paleobiology: the role of cladistic analysis in systematic paleontology. *Paleobiology* 7:456-468.
- Cushing, J. T., C. F. Delaney, and G. M. Gutting (eds.). 1984. Science and Reality: Recent Work in the Philosophy of Science. University of Notre Dame Press, Notre Dame.
- Engelmann, G. F. and E. O. Wiley. 1977. The place of ancestor-descendant relationships in phylogeny reconstruction. *Syst. Zool.* 26:1-11.
- Farris, J. S. 1970. Methods for computing Wagner trees. *Syst. Zool.* 19:83-92.
- Farris, J. S. 1979. The information content of the phylogenetic system. *Syst. Zool.* 28:48-59.
- Farris, J. S. 1983. The logical basis of phylogenetic analysis. In N. I. Platnick and V. A. Funk (eds.), Advances in Cladistics. Columbia University Press, New York, vol. 2, pp. 7-36.
- Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27:401-410.
- Gaffney, E. S. 1979. An introduction to the logic of phylogeny reconstruction. In J. Cracraft and N. Eldredge (eds.), Phylogenetic Analysis and Paleontology. Columbia University Press, New York, pp. 79-112.
- Hanson, N. R. 1958. Patterns of Discovery. Cambridge University Press, Cambridge, MA.
- Hartman, S. E. 1986. A molar perspective on hominoid systematics. *Am. J. Phys. Anth.* 69:212.
- Joysey, K. A. and A. E. Friday (eds.). 1982. Problems of Phylogenetic Reconstruction. Academic Press, New York.
- Kluge, A. G. and J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18:1-32.
- Lambert, K. and G. G. G. Brittan. 1970. An Introduction to the Philosophy of Science. Prentice-Hall, Englewood Cliffs, New Jersey.
- Laudan, L. 1981. Science and Hypothesis: Historical Essays on Scientific Methodology. D. Reidel Publishing Co., Dordrecht, Holland.
- Laudan, L. 1984. A confutation of convergent realism. In J. Leplin (ed.), Scientific Realism. University of California Press, Berkeley.
- Leplin, J. (ed.). Scientific Realism. University of California Press, Berkeley.
- Lewis, O. J. 1964. The homologies of the mammalian tarsal bones. *J. Anat.* 98:195-208.
- Lewis, O. J. 1972. The evolution of the hallucial tarso-metatarsal joint in the Anthropoidea. *Am. J. Phys. Anth.* 37:13-34.

- Martin, L. 1988. Which ape is man's closest kin? *The Sciences* 28(2):52-58.
- Martin, R. D. 1986. Are fruit bats primates? *Nature* 320:482-483.
- Nagel, E. 1961. The Structure of Science: Problems in the Logic of Scientific Explanation. Harcourt, Brace, and World, New York.
- Novacek, M. J. 1980. Cranioskeletal features in tupaiids and selected Eutheria as phylogenetic evidence. In W. P. Luckett (ed.), Comparative Biology and Evolutionary Relationships of Tree Shrews. Plenum Press, New York, pp. 35-93.
- Panchen, A. L. 1982. The use of parsimony in testing phylogenetic hypotheses. *Zool. J. Linn. Soc.* 74:305-328.
- Pettigrew, J. D. 1986. Flying primates? Megabats have the advanced pathway from eye to midbrain. *Science* 231:1304-1306.
- Salmon, W. C. 1966. The Foundations of Scientific Inference. University of Pittsburgh Press, Pittsburgh.
- Schwartz, J. H. 1988. Orang-utans and Human Origins. Houghton Mifflin Company, New York.
- Sober, E. R. 1983. Parsimony methods in systematics. In N. I. Platnick and V. A. Funk (eds.), Advances in Cladistics. Columbia University Press, New York, vol. 2, pp. 37-48.
- Ward, S. C. 1986. Paranasal pneumatization of the frontal bone in early Miocene hominoids. *Am. J. Phys. Anth.* 69:276.
- Watkins, J. 1984. Science and Scepticism. Princeton University Press, Princeton, New Jersey.
- Wikander, R. 1985. Parsimony and testability. *Can. J. Zool.* 63:728-732.
- Wikander, R. 1986. Parsimony and phylogeny reconstruction. *Am. J. Phys. Anth.* 69:279.
- Wikander, R. 1987. Functional anatomy of the hindlimb skeleton of callitrichids. *Am. J. Phys. Anth.* 72:269.
- Wikander, R., H. H. Covert, and D. D. DeBlieux. 1984. The utility of the prehallux in studies of primate phylogeny. *Am. J. Phys. Anth.* 63:234-235.
- Wikander, R., H. H. Covert, and D. D. DeBlieux. 1986. Ontogenetic, intraspecific, and interspecific variation of the prehallux in primates: implications for its utility in the assessment of phylogeny. *Am. J. Phys. Anth.* 70:513-523.