ON DIVERSITY*

Martin L. Weitzman

An oft-repeated goal in many contexts is the "preservation of diversity." But what is the diversity function to be optimized? This paper shows how a reasonable measure of the "value of diversity" of a collection of objects can be recursively generated from more fundamental information about the dissimilarity-distance between any pair of objects in the set. The diversity function is shown to satisfy a basic dynamic programming equation, which in a well-defined sense generates an optimal classification scheme. A surprisingly rich theory of diversity emerges, having ramifications for several disciplines. Implications and applications are discussed.

I. Introduction

"Loss of diversity" is a much-lamented condition nowadays. One sees such a phrase applied loosely in a variety of contexts, including the realms of biological species, landmark buildings, historic sites, languages, artifacts, habitats, even ways of life. Often there is an implicit injunction to preserve diversity because it represents a higher value than other things, which by comparison are "only money." Yet the laws of economics apply to diversity also. We cannot preserve everything. There are no free lunches for diversity. Given our limited resources, preservation of diversity in one context can only be accomplished at some real opportunity cost in terms of well-being forgone in other spheres of life, including, possibly, a loss of diversity somewhere else in the system.

Actual implementation of any injunction to "preserve diversity" is hampered by the lack of an operational framework or objective function. We badly need a more-or-less consistent conceptual framework and a more-or-less usable measure of the value of diversity that can tell us how to trade off one form of diversity against another. The aim of this paper is to provide such a framework and such a measure.

If a value of diversity function can be meaningfully postulated, then it can, at least in principle, be made commensurate with other

benefits and costs, and the general form of the resource allocation problem is in principle well defined. There are presumably some limits on the feasible preservation actions that can be taken, represented by budget constraints or other limitations. Each feasible action induces a probability distribution for what survives, for how long. (There is, perhaps, a significant amount of correlation in survival functions among some species.) The optimal preservation policy may be defined as the feasible action that yields the highest present discounted expected value of diversity (plus whatever other net benefits are attributed to various species). This is in the form of a classical constrained optimization problem. Since the constraint set is in principle well defined, the major unresolved conceptual issue involves defining a meaningful value-of-diversity objective function. The remainder of the paper concentrates on this critical aspect of the problem.

II. THE NATURE OF THE PROBLEM

For convenience and consistency, in what follows I shall employ largely biological metaphors. However, as was pointed out, the mathematical essence of the problem applies to a broader setting and is perhaps more appropriately understood at a higher level of abstraction. The basic underlying unit will be called a "species." A "species" could stand for a genuine species in the traditional biological sense of being a reproductively isolated group having a history of strict genetic divergence from other groups; or, "species" could mean an individual, a subspecies, a specimen, an object, or almost anything else depending on the context. In this paper a species corresponds to what is sometimes called an *operational taxonomic unit* (OTU), broadly defined.

Assume that there is given some very big universe $\tilde{S}$ containing a very large number $\tilde{n}$ of potential species. Of more immediate interest to the decision maker is some set $S \subset \tilde{S}$ containing n member species (written as $|S| = n$). Assume that if i and j are members of S, then some distance or dissimilarity measure between them $d(i,j)$ is given.

The crucial starting point essentially takes as given a cardinal measure of the degree of dissimilarity or difference between any pair of species (i,j), called their distance $d(i,j)$, which satisfies[1]

$$(1) \qquad\qquad d(i,j) \geq 0,$$

---

1. Note that I am not assuming the triangle inequality holds, so that in principle the formulation covers nonmetric distances.

(2) $d(i,i) = 0,$

(3) $d(i,j) = d(j,i).$

The exact origin of the distance measure (1)–(3) is not directly relevant to the purposes of this paper. What matters is that for any pair of species belonging to S there is a nonnegative, symmetric distance measure that meaningfully expresses the difference or dissimilarity between the pair. If $|S| = n$, then altogether there are $n(n - 1)/2$ such pairwise symmetric distance-dissimilarity coefficients, which form the given primitives for the problem of quantifying the diversity of S.

The distance between two species might be derived as a hedonic weighted sum of distances between more fundamental micro-characteristics, so that conceptually $d(i,j)$ represents the weighted number of observable "character-state differences" between species i and j. For architectural examples the micro-characteristics might involve the period of the building, its style, distinguishing features, location, and so forth. In the biological case it might be appropriate to use some aggregate measure of genetic distance based on molecular properties,[2] or, perhaps more ideally for some purposes, time back to the most recent common ancestor. Dissimilarity between two languages might be measured by the fraction of noncognate words. Yet other constructs are possible in other areas.

For the same collection of objects, it might be appropriate to use different distance measures in different contexts depending on the purpose. In other words, the nature of the problem should determine the appropriate distance-dissimilarity measure. Distance is such an absolutely fundamental concept in the measurement of dissimilarity that it must play an essential role in any meaningful theory of diversity or classification. Therefore, it seems to me, the focus of theoretical discussion must be about whether or not a particular set of distances is appropriate for the measurement of pairwise dissimilarity in a particular context, not about whether or not such distances exist in the first place.

The point is not that such pairwise distance-dissimilarity measures as have been described above are unambiguously defined or that they are easy to obtain. The point is rather that if dissimilarity cannot be defined for a pair of objects, then it is

difficult to imagine how collective dissimilarity can be defined for a collection of more than two objects. Thus, it seems to me that if there is to be any hope at all of measuring meaningfully the diversity of a set, and therefore of solving constrained optimization problems involving diversity, it must first be possible to measure the diversity between any two members of the set. (Even then, as we shall see, it is far from trivial to build up from pairwise comparisons to n-tuple comparisons.) In this paper, then, I start from the premise that a meaningful dissimilarity measure (1)–(3) is given for all pairwise comparisons of interest.

The fundamental task of the paper can now be stated in broad terms. Let the value of diversity as a function of set S be denoted $V(S)$.

The basic problem is to construct an appropriate measure of the diversity of a set S, denoted V(S), out of given distance-dissimilarity measures $\{d(i,j)\}$ between all pairs of members $i \in S$, $j \in S$.

There is not really much literature to cite that is directly relevant to the problem at hand,[3] although, because the solution will touch upon many disciplines, there is an enormous literature on related themes. For example, there is an extensive literature on reconstructing evolutionary trees out of more primitive information on genetic distances.[4] In principle, there could be some systematic connections between topological characteristics of classificatory trees, pairwise distances, and diversity. This is actually a major theme of the current paper that will be rigorously explored presently.[5]

3. The most directly relevant antecedent to the present paper is Solow, Polasky, and Broadus [1991]. In this pioneering work the authors give a convincing critique of some simple proposals that have been made for measuring biodiversity and go on to present their own candidate, called a "preservation measure," which they use to analyze a number of conservation issues. In its scope and style, and in the themes and applications it highlights, the approach of Solow et al. served as an inspiration for the present paper. What I think distinguishes this paper from theirs is primarily more emphasis on the theoretical development of an alternative, and in my opinion generally more appropriate, diversity function.
4. For literature on inferring phylogenetic relationships from molecular data, see, e.g., the comprehensive survey of Swofford and Olsen [1990] and the further references they cite. Felsenstein [1988] is another complete source. Li and Grauer [1991] constitutes an excellent overview of the basic issues of molecular evolution and molecular systemics.
5. Thus far, I think it is fair to say that the biological literature has just tentatively begun to grapple with these connections. See Erwin [1991], Krajewski [1991], May [1990], Vane-Wright et al. [1991], and the references cited in these works.

III. DIVERSITY IN A "PERFECT TAXONOMY"

In order to motivate the definition of diversity in the general case, and also because it is of importance in its own right, the diversity function will first be defined for the special case of a "perfect taxonomy."

Let Q be any non-empty proper subset of S:

(4) $$\varnothing \subset Q \subset S.$$

Let j be any element belonging to S but not to Q:

(5) $$j \in S\backslash Q.$$

The standard definition of the distance from the point j to the set Q is

(6) $$d(j,Q) \equiv \min_{i \in Q} d(j,i).$$

The distance $d(j,Q)$ defined by (6) is a measure of the difference, dissimilarity, or diversity between species j and collection Q because it is equal to the difference, dissimilarity, or diversity between j and its "nearest neighbor" or "closest relative" in Q. When $d(j,Q)$ is small, only a little diversity should be added by appending species j to the collection Q, because j is already closely related to at least one of the species in Q. Conversely, when $d(j,Q)$ is big, then a lot of diversity should be added by combining j with Q because species j is only distantly related to any species currently in Q.

Heuristically speaking, $d(j,Q)$ should act somewhat like a derivative or first difference of the diversity function $V(\cdot)$ with respect to j, evaluated at Q. A way of formalizing this intuition might be to study the solution of the following analogue of a differential or first-difference equation:

(7) $$V(Q \cup j) - V(Q) = d(j,Q).$$

Suppose that the analogy with calculus is stretched a bit further. *If* there exists a function $V(\cdot)$ fulfilling the condition (7) for *all* Q and for *all* j, *then* it would seem altogether fitting and natural to call this function a *diversity function*. Furthermore, if such a function is unique (up to an additive constant of integration), it seems legitimate to call $V(\cdot)$ *the* diversity function (up to the additive constant).

Unfortunately, it turns out that (7) cannot hold in the general case for all j and Q satisfying (4), (5). The most that can be hoped

TABLE I
MATRIX OF ULTRAMETRIC DISTANCES

	1	2	3	4	5	6
1	0	159	250	357	532	532
2	159	0	250	357	532	532
3	250	250	0	357	532	532
4	357	357	357	0	532	532
5	532	532	532	532	0	126
6	532	532	532	532	126	0

for in the general case is that (7) should be "close" to holding in some sense.

There is an interesting and important special case in which (7) does hold for all j and Q satisfying (4), (5). This is the special case where all (pairwise) distances are *ultrametric*.

Points belonging to S have ultrametric distances if for any triple $\{i,j,k\} \in S$:

(8) $\max \{d(i,j),d(j,k),d(i,k)\} = \text{mid} \{d(i,j),d(j,k),d(i,k)\}$.

Condition (8) means that for the three possible pairwise distances between any three points, the two greatest distances are equal.

A numerical example of ultrametric distances among six species is given in Table I. Just to make things interesting, it happens to be the case that the numbers in Table I are based on real-world data about six major species of higher primates, one of which is the humans.[6] The names of the six hominoid species are given in Table II.[7]

Now it turns out that ultrametric distances have a very important geometric property. All of the information they contain can be represented graphically by a hierarchical or taxonomic tree. Such a tree is drawn in Figure I for the distances that correspond to Table I. (Technically speaking, what is depicted in Figure I is a rooted directed tree, but I shall refer to it here simply as a tree.)

6. Where it is appropriate, I shall later indicate how the distances are formed. I have omitted the pygmy chimpanzee (*Pan paniscus*) because its inclusion would muddy some theoretical points I want to make clearly in the example, and also because the genetic difference between common chimps and pygmy chimps is no greater than between different subspecies of, e.g., orangutans.
7. Actually, the identity of the six species is intended only as an attention-grabbing ploy to sharpen the reader's possible interest in what is essentially a numerical example; in the paper no further reference will be made to the significance of the numbers themselves.

TABLE II
SPECIES LIST FOR THE EXAMPLE

Number	Common name	Scientific name
1	human	*Homo sapiens*
2	common chimpanzee	*Pan troglodytes*
3	gorilla	*Gorilla gorilla*
4	orangutan	*Pongo pygmaeus*
5	common gibbon	*Hylobates lar*
6	siamang gibbon	*Hylobates syndactylus*

The six existing species are depicted as six twig-tip terminal nodes. There are five interior ancestor nodes. The distance between any two species of a tree is depicted as the vertical distance back to their nearest common ancestor node.

The fact that ultrametric distances can be represented as a tree and that any tree defines ultrametric distances will be taken here as a mathematical truth.[8] Ultrametric distances are identified with a "perfect taxonomy" for the following reason.

Essentially, it is being assumed that there exists a perfectly complete genealogical tree, as represented by a diagram like Figure I, showing how the existing species of today evolved from a common ancestor. It is as if the fossil record is completely preserved and accurately dated: all times of branching and first appearances of relevant species are known exactly.

A "clade" is a subgroup of species that each share a particular common ancestor.[9] For the tree of Figure I the species sets {1,2}, {1,2,3}, {1,2,3,4}, and {5,6} are clades. Within a perfect taxonomy the evolutionary tree forms a series of perfectly nested clades. In this context of the paleontologist's dream phylogeny, it is natural to define the distance between any two species as the time ago when they diverged from a common ancestor. This natural, and for some situations ideal, distance measure—time removed from the most recent common ancestor—possesses some remarkable, intuitively very satisfying, properties deriving from ultrametricity.

It is not difficult to show that in a perfect taxonomy, and only

8. Actually, it is difficult for me to find in the literature a mathematically rigorous, comprehensive treatment of the relationship between ultrametric distances and rooted directed trees.

9. In terminology that may be more familiar to economists, think of an evolutionary tree as a game tree. A clade corresponds exactly to all the terminal nodes of some proper subgame of the full game. I am indebted to Avinash Dixit for pointing out this analogy.
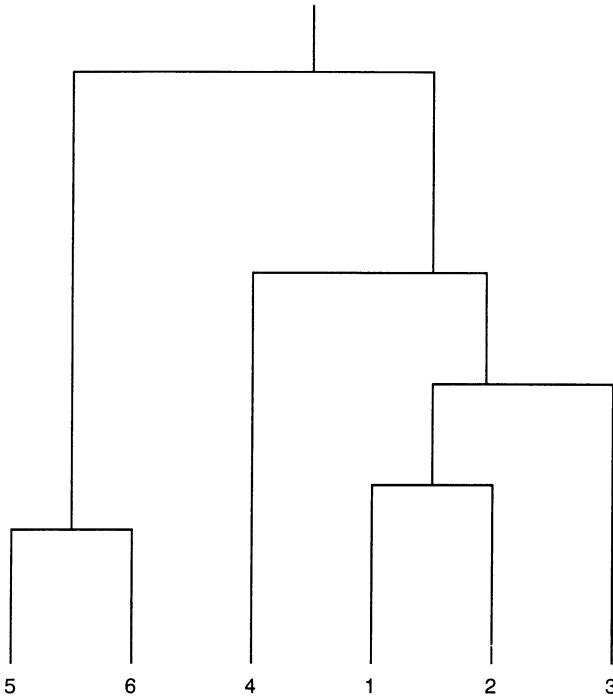
FIGURE I
The Maximum Likelihood Tree Representation

in a perfect taxonomy, condition (7) holds for *all j* and *all Q* satisfying (4), (5). Furthermore, there is a strikingly simple geometric interpretation of the diversity function in this special case.

Let the *length* of an evolutionary tree, like what is depicted in Figure I, be defined to be the total lengths of all its vertical branches (including the branch of the common ancestor of the entire family back to some unspecified outgroup). Then, up to an arbitrary additive constant, the diversity function defined by (7) (holding for *all j* and *all Q* satisfying (4), (5)) is equal to the summed total branch length of the perfect taxonomy tree. Essentially, *with ultrametric distances, the diversity function is the length of the associated taxonomic tree.*

This important statement should be intuitively plausible upon reflection. A rigorous proof is not difficult, but is surprisingly lengthy and symbol-laden, primarily because it requires some

not-inconsiderable setup effort to define rigorously in abstract mathematical notation certain technical properties of rooted directed tree structures and subtree structures.[10] Here only a sketch of the proof of this basic proposition is provided.

The theorem to be proved is the following.

Let S be any set of species possessing the ultrametric distance property (8). Strictly for normalization purposes, suppose that there exists some, possibly hypothetical, zeroth species, called an *outgroup* species, having the normalization property,

$$d(i,0) \equiv d_0, \qquad \forall i \in S,$$

where, for convenience,

$$d_0 > d(i,j), \qquad \forall i \in S, \quad \forall j \in S.$$

Let Q be any subset of S (Q satisfies (4)). Then it is immediate that the species of Q also possess the ultrametric distance property (8). Let T_Q represent the perfect taxonomy tree associated with Q. Let the total (vertical) branch length of the tree T_Q (including the branch back to the ancestor of the outgroup species zero) be denoted

$$L(Q).$$

THEOREM. With ultrametric distances there exists a function $V(Q)$ consistently fulfilling the condition (7) for *all* j and *all* Q satisfying (4), (5). The function $V(Q)$ can be expressed as

(9) $$V(Q) = L(Q) + K,$$

where K is any constant.

Proof. For the reasons cited above, I present here only an outline of the proof. The interested reader should not have much trouble filling in the details, although, as indicated, they are time- and space-consuming.

The heart of the proof proceeds by induction on the size of S. Suppose that the theorem holds for all S with $|S| = n$. (This is readily confirmed for $n = 2$.) Now let S be any set with $|S| = n + 1$.

Let Q be any subset of S. For $|Q| \leq n$, the theorem is holding by induction. We now show constructively how to extend the $V(Q)$ function defined by the induction assumption for $|Q| \leq n$ to define $V(S)$ consistent with the statement of the theorem.

10. See Weitzman [1991a] for a sense of what is involved.

Because ultrametric distances can be represented as a tree (which we are taking as a mathematical fact), and because any subtree of a tree is also a tree (again taken as a mathematical fact), we have

(10)
$$L(Q \cup j) - L(Q) = d(j,Q), \quad \forall Q \subset S, \quad \forall j \in S \backslash Q.$$

From the induction assumption (9) applied to $Q \equiv S \backslash i$, we have

(11) $$V(S \backslash i) = L(S \backslash i) + K, \quad \forall i \in S,$$

where K is chosen to be the *same* arbitrary constant for each i belonging to S.

Combining (11) with the induction assumption, we have

(12) $$V(Q) = L(Q) + K, \quad \forall Q \subset S.$$

Now define $V(S)$ as follows:

(13) $$V(S) \equiv L(S) + K.$$

Combining (10), (12), and (13), we have

(14) $$V(Q \cup j) - V(Q) = d(j,Q), \quad \forall Q \subset S, \quad \forall j \in S \backslash Q,$$
(15) $$V(Q) = L(Q) + K, \quad \forall Q \subseteq S.$$

Conditions (14), (15) complete the induction argument.|

The above result shows that ultrametric distances provide something akin to an integrability condition, allowing, in this special case, some analogue of the fundamental theorem of calculus to hold for all existing species. When any species becomes extinct, the loss of diversity equals the species' distance from its closest relative, and this myopic formula can be repeated indefinitely over any extinction pattern, because any sub-evolutionary tree of an evolutionary tree is also an evolutionary tree. When a species becomes extinct, the loss of diversity is calculated as if its evolutionary branch were snapped off the rest of the tree and discarded. This sharp mental image, properly used, permits a quick, exact visualization of the effects of various combinations of species losses on diversity in the special case of perfect taxonomy.

A simple example may help to illustrate the basic issues. In Figure I is depicted a family tree representing the evolutionary history of six existing species. The two most closely related species are 5 and 6, so that the smallest loss from extinction of a single species occurs if one of these two vanish. However, an analytical preservationist must be careful here. If, after species 5 goes extinct,

species 6 also goes extinct, then the overall loss could be catastrophic since a whole evolutionary line will have been wiped out. While the diversity loss of 5 or of 6 is lower than that of any other single species in the collection, the diversity loss of the pair (5,6) is greater than the diversity loss of any other pair in the set. Hence, an optimal conservation strategy might be to concentrate relatively few resources on saving species 5, if species 6 is reasonably safe, or it might involve concentrating relatively large resources on saving species 5, if species 6 has a high danger of extinction. I hope this kind of example, which could be repeated over a wide variety of different situations, illustrates the power of using the simple geometric interpretation of (14), (15) as a conceptual aid for analyzing policy options concerning preservation of diversity.

The perfect taxonomy structure induced by ultrametric distances allows some other powerful insights into the form of an optimal conservation policy. Consider, for example, the following idealized situation involving sharply posed preservation issues in a context of ultrametric distances.

Let the set S consist of n species denoted by $i = 1, 2, \ldots, n$. Let the (independent) probability that species i survives be denoted x_i. Each column n-vector $X \equiv (x_i)$ of survival probabilities defines an expected diversity function:

$$U(X) \equiv E_X(V).$$

Suppose that the objective function is of the form,

$$\Phi(X) = BX + U(X),$$

where b_i is the direct net benefit of species i and $B \equiv (b_i)$ is the row n-vector of direct net benefit coefficients.

Suppose that the cost of preserving species i with probability x_i is equal to $c_i x_i$. Let the row n-vector of cost coefficients be $C \equiv (c_i)$. Let the total preservation budget be M.

The simplest form of a constrained expected diversity-maximizing problem might be formulated as

(16)   maximize $\Phi(X)$

subject to

(17) $CX \leq M$,

(18) $0 \leq x_i \leq 1$,   for $i = 1, 2, \ldots, n$.

The above constrained optimization problem is well defined, but it looks like a combinatoric nightmare. Actually, in the case of

ultrametric distances a simple myopic algorithm is available for solving (16)–(18). It is here stated without proof.

The algorithm proceeds by eliminating the least valuable species, one species at a time, until the budget constraint is just met.

Suppose that at some iteration the subset $Q \subseteq S$ of species exists with probability one, while the subset $S \backslash Q$ of species is extinct (exists with probability zero). Suppose that the budget constraint is not being met:

$$\sum_{i \in Q} c_i x_i > M.$$

The next step is to find the relatively least desirable species of Q. This is the species $j(Q) \in Q$ that satisfies the condition,

$$\frac{b_j + d(j, Q \backslash j)}{c_j} = \min_{i \in Q} \left\{ \frac{b_i + d(i, Q \backslash i)}{c_i} \right\}.$$

The probability x_j is then brought down continuously from one toward zero until either the budget constraint is met or species j is eliminated, whichever occurs first. In the latter case, a new species set Q is defined which is equal to the previous species set Q minus the species j. (Note: this will change some of the remaining $\{d(i, Q \backslash i)\}$ coefficients.) The procedure is repeated until the budget constraint is just met, at which point the algorithm has converged. The relevant theorem (not proved here) is that in the case of ultrametric distances such a myopic algorithm yields an optimal policy in the sense of satisfying (16)–(18). The theorem justifies using at each iteration a myopic benefit-cost ratio consisting of the traditional ratio of direct benefits to costs plus the diversity loss per preservation dollar.

The import of this approach consists in giving a rigorous global significance to the strictly local decision-making index of species diversity loss per unit of conservation resources. Comparing "expected diversity loss per preservation dollar" among species thus turns out to be a legitimate extension of cost-benefit analysis.[11]

We turn now to the main theme of this paper, which is the meaning and significance of diversity in the extremely general situation where no conditions other than (1), (2), and (3) are placed

---

11. This theme is generalized and applied to an actual numerical example (drawn from crane data) in Weitzman [1991b].

on the pairwise primitives of dissimilarity-distance. It is important to have a general theory of diversity because, as I have argued, the nature and context of the problem should determine the appropriate dissimilarity concept, which leaves a wide scope for possible distance measures.

IV. DIVERSITY IN THE GENERAL CASE

It is only rarely that distances come in ultrametric form as shown in the example of Table I. Far more frequently, distances are not ultrametric. For example, distances might be given as shown in Table III, which displays the actual raw data that was transformed into the numbers of Table I.[12] How the transformation from the non-ultrametric distances of Table III to the "as if" ultrametric distances of Table I occurred will be explained presently, for this actually represents a central theme of the present paper. In any event we are typically confronted initially with a situation where the given pairwise symmetric dissimilarity-distance measures are not ultrametric.

In the general case of arbitrary distances, *the diversity function V(S) is inductively defined to be the solution of the recursion:*

$$(19) \qquad V(S) = \max_{i \in S} \{V(S\backslash i) + d(i,S\backslash i)\}.$$

The dynamic programming equation (19) is the centerpiece of the present paper.[13] The solution of (19) is unique once the initial conditions,

$$(20) \qquad V(i) \equiv d_0, \qquad \forall i,$$

are specified for any d_0. (Depending upon the particular application, it is typically most convenient to normalize d_0 by setting it equal either to zero or to some large constant.)

There are several possible axiomatic approaches that can be used to justify the diversity function (19). These axiomatic treatments are suggestively motivating, as I hope to indicate. However,

12. The numbers in Table III come straight out of Table 4 in Caccone and Powell [1989]. I have omitted the nonhominoid baboon and the almost conspecific pygmy chimpanzee primarily to be able to illustrate some theoretical points more crisply.

13. The genesis of the present paper really began with equation (19). I am grateful to Sergiu Hart for starting me going in this direction by proposing the length of the minimal spanning tree as a measure of diversity. Focusing on what was wrong with the minimal spanning tree as a diversity indicator led me to invent the dynamic programming function (19) as a solution concept, from which the rest of the ideas developed.

TABLE III
MATRIX OF GIVEN DISTANCES

	1	2	3	4	5	6
1	0	159	250	349	495	513
2	159	0	234	328	448	448
3	250	234	0	357	532	498
4	349	328	357	0	477	488
5	495	448	532	477	0	126
6	513	448	498	488	126	0

the real argument for the diversity measure being proposed here is that it "works"—in the sense of creating a useful and consistent conceptual framework, while other measures "do not work"—in the sense that they violate one or more essential properties that a plausible diversity function should possess. The force of this point can only be appreciated by actually trying out some other candidates for a value of diversity function and seeing at first hand where they fail to make sense.[14]

We are faced, in the general case, with a situation where (7) does not hold for all Q and for all j defined by (4), (5). The natural instinct is to try to make (7) hold as nearly as possible under the circumstances.

The following condition seems like a basic axiom that is reasonable to impose on any diversity function.

MONOTONICITY IN SPECIES. If species j is added to collection Q, then

$$(21) \qquad V(Q \cup j) \geq V(Q) + d(j,Q), \qquad \forall Q, \quad \forall j \notin Q,$$

where $d(j,Q)$ is the familiar (minimal) distance from point j to set Q defined by (6).

The monotonicity in species condition (21) expresses the intuitively desirable idea that the addition of any species to a group of species should increase diversity by at least the dissimilarity of that species from its closest relative among the already existing

14. There is simply not enough space in this article to go through the many other seemingly reasonable definitions of diversity to show, in a series of convincing examples, where they are seriously flawed, even though, in the end, superiority over the alternatives is the most telling argument in favor of selecting (19) as a diversity function. The reader seriously interested in the subject should definitely try this exercise.

group of species. Or, conversely, monotonicity in species means that the extinction of any species of an ensemble causes a decline in diversity by no less than the distance of the extinguished species from its nearest neighbor in the ensemble.

Monotonicity in species is a "loose" property in the sense that it does not at all define a unique function because many diversity measures can be made to satisfy the inequality (21). There are at least two ways to add a supplementary condition that would make the inequality (21) hold so "tight" that it yields, in effect, the dynamic programming equation (19).

The first approach is the most direct. View condition (21) as a (potentially very large) set of constraints that must hold for *all* Q and for *all* j. Impose the uniform initializing condition (20). Then simply define the diversity of S to be the minimum possible $V(S)$ that satisfies (21), (20).

The reason that this direct approach yields the dynamic programming recursion (19) is as follows. Suppose, by induction, that the diversity functions $\{V(S\backslash i)\}$ have been defined for all i belonging to S. Then the smallest possible value for the diversity of S that would be consistent with (21) must satisfy the condition,

$$(22) \qquad V(S) \equiv minimum\, V,$$

subject to

$$(23) \qquad V \geq V(S\backslash i) + d(i,S\backslash i), \qquad \forall i \in S.$$

It is straightforward to confirm that the solution of (22), (23) is (19), which both proves the assertion and continues the induction argument to the next stage.

The problem of finding the smallest possible diversity function consistent with (21) can be recast as an insightful evolutionary metaphor.

In this interpretation the distance $d(i,j)$ stands for the number of (possibly weighted) character-state differences between i and j. For any set Q of existing species, $V(Q)$ here stands for the evolutionary length of Q, meaning the total number of character-state changes required to explain the evolution of Q under some branching representation of the evolutionary process. Suppose that species $j \notin Q$ is added to Q to form the new set $Q \cup j$ of existing species. The number of *extra* character-state changes required to explain the evolution of j is at least the difference in character-state changes between j and its closest relative in Q, which by (6) is $d(j,Q)$. If j is added to Q, then at least $d(j,Q)$ additional

character-state changes need to be explained. Therefore, any properly scaled feasible measure of evolutionary distance should simultaneously satisfy, for *all Q* and for *all j*, the basic consistency conditions (21).

It seems natural to define the diversity of S, denoted $V(S)$, to be the length of the tightest or most parsimonious feasible reconstruction of S, in the sense of being the minimal number of character-state changes required to account for the evolution of S. By the same argument as before, $V(S)$ so defined must satisfy (22), (23), and, by extension, (19). Thus, equation (19) has the interpretation of generating the most parsimonious "minimal evolution" branching structure that gives rise to the species of S.

A second way of forcing the inequality (21) to hold so "tight" that it yields the dynamic programming equation (19) is to add an extra axiom to (21) called the "link property." This new condition can be stated as follows.

LINK PROPERTY. For all S, $|S| \geq 2$, there exists at least one species $j(S) \in S$, called the "link" species, that satisfies

$$(24) \qquad V(S) = d(j, S \backslash j) + V(S \backslash j).$$

As was shown in the last section, an especially appealing theoretical structure emerges in the case of ultrametric distances, where, in effect, (24) holds for *all* $j \in S$. Unfortunately, it is mathematically impossible that (24) can be true for all $j \in S$ in the general case. But from the link property it will at least always be true that the elimination of *some* species $j(S)$ will reduce diversity by exactly the distance of that species from its closest relative. The link property provides at least one tight natural connection between the derived value of diversity measure for any set and the primary distance data on which it is based.

That conditions (21) and (24) imply condition (19) is a fairly straightforward argument.

There is also a probabilistic way of motivating the basic dynamic programming recursion (19) that deserves to be treated here.

One of the most commonly cited reasons for maximizing expected biodiversity is to maintain a kind of natural "portfolio diversification" of future options for finding new sources of food, medicine, and so forth.[15]

---

15. For some background motivation see Oldfield [1984].

Suppose, for concreteness, that we are speaking of finding a pharmacological cure for some disease. If a species contains a cure, it will only become revealed over time, in the future. Thus, when a species becomes extinct, the chance is lost forever that the species may be of later help in providing a medicine for treating the disease. What should we be preserving in such a context?

In this interpretation let

$$(25) \qquad\qquad P(i,j)$$

stand for an upper bound on the probability that species i does *not* contain a cure for the disease, given that species j does *not* contain a cure.

The data represented by (25) are the basic, given, reduced-form primitives of the model.

It is assumed that the given conditional probability coefficients (25) are symmetric;[16] for all i and j belonging to S,

$$P(i,j) = P(j,i).$$

In what follows, assume any particular evolutionary branching tree structure T out of all possible rooted directed trees that yield the species of S as labeled twig-tip end-nodes of the evolutionary process. Each possible tree T defines a set of ancestor interior-nodes $A_T(S)$ ($n - 1$ of them in the bifurcating case), located within the branching structure.

Think of evolution as a branching process that results in the accretion of many tiny boxes. When two new species bifurcate from an ancestor node, they keep all the same tiny boxes they shared in common to that point, but henceforth they begin independently accreting different tiny boxes. If there is a pharmacological treatment for the disease, it will be found in one of the tiny boxes that have been accumulated along the evolutionary tree. In this model the key structural assumption is that once a cure is contained in a parent node, then it is fixed or locked into all of the subsequent offspring nodes.

All statements that follow are with respect to the particular branching structure T being assumed. In other words, for convenience we are dropping the subscript T from the notation that follows, understanding that it is implicitly there.

16. Symmetry of the conditional probability coefficients (25) is equivalent to postulating that the unconditional probabilities $P(i)$ and $P(j)$ are equal, which seems like an appropriate symmetry assumption for species at the same general level of complexity.

Let Q and j satisfy (4), (5). Let

$$P(j|Q)$$

be the conditional probability that species j does *not* contain a cure given that each species of the set Q does *not* contain a cure.

Let

$$A(Q)$$

represent the set of all ancestor nodes of the set of species Q.

Then,

$$(26) \qquad P(j|Q) = P(j|A(Q)).$$

(The only way that the fact that Q does not contain a cure transmits information relevant to whether or not j contains a cure is through the knowledge that the ancestor nodes $A(Q)$ could not have contained a cure.)

Let

$$a(j,Q)$$

stand for the most immediate ancestor of j in the set $A(Q)$.

Then,

$$(27) \qquad P(j|A(Q)) = P(j|a(j,Q)).$$

(The entire relevance for the probability that j does not contain a cure given that $A(Q)$ does not contain a cure is summarized by the information that $a(j,Q)$, the most recent ancestor of j in $A(Q)$, does not contain a cure.)

Applying basic probability theory to this special structure,

$$(28) \qquad P(j|i) = P(j|a(j,Q)) \cdot P(a(j,Q)|i) \qquad \forall i \in Q.$$

Taking the maximum of both sides of expression (28) over all $i \in Q$ yields

$$(29) \qquad \max_{i \in Q} P(j|i) = P(j|a(j,Q)) \cdot \max_{i \in Q} P(a(j,Q)|i).$$

Now $a(j,Q)$ $(\in A(Q))$ must be an ancestor node for *some* (at least one) $k \in Q$, implying that

$$(30) \qquad P(a(j,Q)|k) = 1 \qquad \text{for some } k \in Q.$$

But then (30) implies that

$$(31) \qquad \max_{i \in Q} P(a(j,Q)|i) = 1.$$

Combining (31) with (29) with (27) with (26), and using definition (25) yields

(32) $$P(j|Q) \leq P(j,Q),$$

where

(33) $$P(j,Q) \equiv \max_{i \in Q} P(j,i).$$

By definition

(34) $$P(j|Q) \equiv P(Q \cup j)/P(Q),$$

where $P(Q)$ stands for the probability that none of the species of Q contain a cure for the disease.

Combining (32) with (34) yields

(35) $$P(Q \cup j) \leq P(Q) \cdot P(j,Q) \qquad \forall Q \subset S, \quad \forall j \in S \backslash Q.$$

Let $\Pi(S)$ stand for the *maximum* value of $P(S)$ under the set of constraints signified by (35):

(36) $$\Pi(S) \equiv \max P(S), \qquad \text{subject to (35).}$$

If $\Pi(S \backslash i)$ were known for all $i \in S$, then (35) (with $j \equiv i$, $Q \equiv S \backslash i$) implies that $\Pi(S)$ defined by (36) must satisfy the dynamic programming recursion,

(37) $$\Pi(S) = \min_{i \in S} \{\Pi(S \backslash i) \cdot P(i, S \backslash i)\}.$$

It is convenient to transform (37) into an equivalent dynamic programming equation that is additive in distances by taking negative logarithms of all probabilities.

Let

(38) $$V(S) \equiv -\log \Pi(S),$$

(39) $$d(i,j) \equiv -\log P(i,j).$$

Combining (33) with (6) with (39) yields

(40) $$d(i, S \backslash i) = -\log P(i, S \backslash i).$$

Using (38), (39), and (40), equation (37) becomes transformed into the equivalent dynamic programming equation (19).

From the isomorphism of (37) with (19), the following important conclusion emerges. When the diversity function defined by (19) is being maximized, there is a well-defined sense in which the worst-case probability of not being able to find a cure for the

disease is simultaneously being minimized. Thus, provided that the distances are appropriately defined, maximizing diversity is equivalent to minimizing the worst-case risk of not being able to avoid some bad outcome in a portfolio choice problem.

It is worthwhile reviewing the essential abstract structure of the argument just completed, because it can be used to motivate another probabilistic interpretation of the fundamental dynamic programming equation (19)—as a condition defining the most likely branching structure that might explain the evolution of the species of S from a common ancestor.

Basically, it has been assumed that there is some property (e.g., the property of *not* containing a cure for a particular disease) with the following essential feature: *if the child-node in an evolutionary tree has the property, then the parent-node also has the property.*

Now the structural property referenced in the above statement can be given several possible interpretations, of which the examples of a (lack of a) pharmacological cure for a disease or a (lack of a) new source of food were just particular applications. Next, suppose that the "structural property" is the property of evolutionary existence itself. A child species-node can exist in an evolutionary tree only if the parent species-node also exists.

Under this interpretation, $P(i,j)$ stands for the a priori probability that species i exists given that species j exists. The symmetric conditional probability coefficients $\{P(i,j)\}$ are taken as given primitives of the problem related to time-distances $\{d(i,j)\}$ by the transformation (39). (The transformation (39) could be given a rigorous basis when all species survival probabilities are exponentially distributed with the same mortality parameter.) In slightly different words, $P(i,j)$ is an a priori estimate of the probability that species i successfully "made it" through the hazardous maze of the evolutionary process to exist at the current time, given the information that species j successfully "made it." Of course, both i and j are actually alive today, but the whole point of the maximum likelihood exercise is to find the branching topology that would explain the fact that all species of S survived—which did not have to be—with higher probability than any other branching topology.[17]

---

17. As $n = |S|$ becomes large, the number of possible rooted trees whose twig tips are the members of S becomes truly enormous, being essentially proportional to n^n in the limit. (See Felsenstein [1978].) There are about 10^{24} possible trees for just twenty species. Numbers like these typically doom attempts to find the maximum

In this "maximum evolutionary likelihood" interpretation, the "property" under discussion is "evolutionary existence." $P(Q)$ stands for the probability that Q exists (meaning that all the species of Q successfully evolved). $P(j|Q)$ stands for the probability that j exists (successfully evolved) given that Q exists.

Now the problem of finding the branching topology that would explain the successful evolution of S with higher probability than any other branching topology has exactly the same mathematical structure as the problem previously treated. Picking up the strand of argument at equation (37), we continue it under the "maximum evolutionary likelihood" interpretation as follows.

Note that although, in the course of the derivation, we implicitly assumed a particular branching structure T, the argument yielded a general sufficient condition (37) seemingly independent of T. Let

$$P_T(S)$$

stand for the probability that none of the n species of S became extinct under T; i.e., all of S "made it" to current existence through the Markov evolutionary maze represented by the particular rooted directed tree T. Then we know from the definition (36) of $\Pi(S)$ that

$$\Pi(S) \geq P_T(S).$$

Therefore, *if* we can constructively exhibit a rooted directed tree T^* with the property that

$$P_{T^*}(S) = \Pi(S),$$

then T^* is the maximum likelihood evolutionary tree of S. The formal statement is that such a T^* would be the evolutionary branching structure that has the highest probability of causing all the species of S to be alive today, out of every possible directed Markovian evolutionary branching structure consistent with (32), (33) holding for *all Q* and *all j*.

It turns out that the maximum likelihood tree of S, here denoted $T^*(S)$, is automatically generated by the solution of the dynamic programming equation (37), which is equivalent to (19). This statement is a variant of what is called the "fundamental representation theorem," to which we turn next.

---

likelihood tree with even moderately sized species sets. The approach presented in this paper typically calculates the global maximum likelihood tree for twenty species in just a few seconds of running time on a good quality personal computer.

To summarize this last line of thought, the basic dynamic programming equation (19) can be interpreted as generating a rooted directed tree. The particular tree that (19) generates can be given a rigorous interpretation of being the maximum likelihood evolutionary branching structure consistent with the given conditional existence probabilities $\{P(i,j)\}$ defined by (39).[18]

V. THE FUNDAMENTAL REPRESENTATION THEOREM

The fundamental representation theorem is essentially a technical statement about the form of an optimal solution to (19). The theorem states that the argmax solution of (19)—i.e., the "link" species $j(S)$ of (24)—must be one of the two closest species in S. The interpretation of this theorem allows us to represent the solution of (19) as a rooted directed tree whose sum of branch lengths, appropriately normalized, equals the diversity function.

For $|S| \geq 2$, define the pair $(g(S), h(S))$ to be "closest relative" functions: $S \to S$ satisfying $g(S) \in S$, $h(S) \in S$, $g(S) \neq h(S)$, and

$$(41) \qquad d(g(S),h(S)) = \underset{\substack{i,j \in S \\ i \neq j}}{\text{minimum}}\ d(i,j).$$

(It does not matter how the pair $(g(S),h(S))$ is ordered or whether it is unique.)

Define the function $W(S)$ as follows. For $|S| = 1$, $W(S) = d_0$. For $|S| \geq 2$, $W(S)$ is the solution of the dynamic programming recursion,

$$(42) \quad W(S) = d(g(S),h(S)) + \max\ \{W(S\setminus g(S)),W(S\setminus h(S))\}.$$

THEOREM (FUNDAMENTAL REPRESENTATION THEOREM). For all S,

$$(43) \qquad\qquad V(S) = W(S).$$

Proof. From (19) it follows that $|S| = 1$ implies that $V(S) = d_0$. For $|S| \geq 2$, let $j(S)$ be any argmax of (19), meaning $j(S) \in S$ is the "link" species of (24) satisfying

$$(44) \qquad V(S) = d(j(S),S\setminus j(S)) + V(S\setminus j(S)).$$

Suppose the *induction assumption*: for all $|S| = n$,

(45) $\quad V(S) = d(g(S),h(S)) + \max \{V(S \backslash g(S)),V(S \backslash h(S))\}.$

Clearly (45) is true for $n = 2$.

Now let $|S| = n + 1$. Suppose that (45) is *not* true for all $|S| = n + 1$. Suppose *per contra* that, for some $|S| = n + 1$,

(46) $\qquad\qquad g(S) \neq j(S) \text{ and } h(S) \neq j(S),$

and

(47) $\quad V(S) > d(g(S),h(S)) + \max \{V(S \backslash g(S)),V(S \backslash h(S))\}.$

From condition (46) and definition (41), $g(S)$ and $h(S)$ must be "closest relatives" in the set $S \backslash j(S)$ (as well as the set S), meaning

(48) $\qquad d(g(S) \backslash j(S)),h(S \backslash j(S))) = d(g(S),h(S)).$

Applying the induction assumption (45) to the set $S \backslash j(S)$, and simplifying by using (48), it follows that

(49) $\qquad\qquad V(S \backslash j(S)) = d(g(S),h(S))$

$\qquad + \max \{V(S \backslash j(S) \backslash g(S)),V(S \backslash j(S) \backslash h(S))\}.$

Substituting (49) into (44), we obtain

(50) $\quad V(S) = d(j(S),S \backslash j(S)) + d(g(S),h(S))$

$\qquad\qquad + \max \{V(S \backslash j(S) \backslash g(S)),V(S \backslash j(S) \backslash h(S))\}.$

Applying (19) to the sets $S \backslash g(S)$ and $S \backslash h(S)$ yields, respectively,

(51) $\qquad V(S \backslash g(S)) \geq d(j(S),S \backslash g(S) \backslash j(S)) + V(S \backslash g(S) \backslash j(S)),$

(52) $\qquad V(S \backslash h(S)) \geq d(j(S),S \backslash h(S) \backslash j(S)) + V(S \backslash h(S) \backslash j(S)).$

Furthermore, from first principles in the definition (37) of the distance from a point to a set,

(53) $\qquad d(j(S),S \backslash g(S) \backslash j(S)) \geq d(j(S),S \backslash j(S)),$

$\qquad\qquad d(j(S),S \backslash h(S) \backslash j(S)) \geq d(j(S),S \backslash j(S)).$

(54) $\quad \max \{V(S \backslash g(S)),V(S \backslash h(S))\}$

$\qquad \geq d(j(S),S \backslash j(S)) + \max \{V(S \backslash g(S) \backslash j(S)),V(S \backslash h(S) \backslash j(S))\}.$

Combining (51), (52), (53), and (54) gives

(55) $\qquad\qquad\qquad V(S) > V(S),$

Using (55) to compare (47) with (50), we have, finally, a contradiction.

The conclusion must be that the assumption (47) is wrong, i.e., for any set S with $|S| = n + 1$,

$$(56) \quad V(S) = d(g(S),h(S)) + \max \left\{V(S\backslash g(S)),V(S\backslash h(S))\right\},$$

the desired induction step.

Since equations (56) and (42) are identical, and since they share the same initial conditions, the unique solutions $V(S)$ and $W(S)$ must also be identical.|

The dynamic programming recursion (42) that defines $W(S)$ for $|S| = n$ involves (at most) 2^n calculations. In this paper I shall not explore in detail the purely numerical aspects of how best to solve or approximate (42). Suffice it to say that a "clever" application of branch and bound methods seems to reduce the number of calculations to about $(1.65)^n$ and that problems of size up to about $n = 35$ can be routinely solved on a good quality personal computer. A mainframe computer could tackle a problem of much larger size.

The basic representation theorem allows or encourages us to think of diversity $V(S)$ from a relatively "difficult" dynamic programming equation (19) (of magnitude n^n) as the solution $W(S)$ of a relatively "easy" dynamic programming equation (42) (of magnitude 2^n). Henceforth, without ambiguity, we can use the two concepts interchangeably and drop the notation $W(S)$ in favor of (the now more inclusive) $V(S)$.

Equation (56) is a dynamic programming recursion which can be interpreted, at any iteration, as a best method of "clustering" the two closest points. Geometrically, the process will define a rooted directed tree, whose twig tips represent the existing species. The fundamental representation theorem generates a genealogical tree describing the hypothetical evolution of S from a hypothetical common ancestor.

It should be appreciated that none of this is in the least bit obvious. That the diversity function (19) should generally be represented by some unrooted spanning tree (what is sometimes called a "network") is perhaps not surprising. What is quite surprising is the fundamental representation theorem with its rooted genealogical tree interpretation.

We show the rooted directed tree interpretation as follows.

Without significant loss of generality, suppose for concreteness that point $g(S)$ uniquely satisfies the maximum in (56). That

is,

(57) $\qquad V(S) = V(S \backslash g(S)) + d(g(S),h(S)),$

meaning that $g(S)$ is the "link" species $j(S)$ of equation (24).

In order to continue the recursion process, from (57) we need only to know the argmax of (56). (This observation has computational significance because any shortcut approximation can be used that tells us which is the larger of $V(S \backslash g(S))$ and $V(S \backslash h(S))$.) In other words, we can think geometrically in terms of "clustering" the points $g(S)$ and $h(S)$ by "casting out" the "link" species $g(S)$ while retaining the "representative" species $h(S)$ to "represent" the pair $(g(S),h(S))$. (The algorithm thus automatically designates a member species to "represent" each clade; this "representative" species embodies, in extremis, the characteristics most distinguishing the given clade from other clades at that taxonomic level because loss of diversity from subtracting the "representative" species is greater than loss of diversity from subtracting the "link" species.) If S was of size n, there is now a dynamic programming problem of dimension $(n - 1)$, because the shortest branch distance $d(g(S),h(S))$ has been eliminated from the evolutionary tree and added to the running sum kept for tabulating V.

The above-described procedure may then be continued in like fashion by similarly clustering the two nearest remaining neighbors of $S \backslash g(S)$, while adding their distance to the running sum tabulation of V, and so on. Each clustering operation reduces the size of the remaining set by one, while adding one term, representing the length of the currently-shortest-distance branch, to the running tabulation of V. The iterations proceed until just one species remains, at which point the running sum of distances is equal to $V(S) - d_0$, where S is the original starting set and d_0 is the normalization constant.

As described above, the algorithm implicit in recursion (56) is a form of pair-group clustering, which is one of the most widely used methods for estimating phylogenies on the basis of distance data. Typically, the artificial distance from each newly created cluster to each other cluster is some weighted average of actual distances. As has been pointed out in the literature, the traditional method can in principle suffer serious discontinuities,[19] and it is ultimately heuristic, lacking any well-defined optimality criterion.[20] The present algorithm has a rigorous foundation—the only one avail-

---

19. See Jardine and Sibson [1971], section 7.4.
20. See Swofford and Olsen [1990] and Felsenstein [1988].

able for a clustering method to my knowledge—but it is less easy to calculate than traditional methods. Optimal clustering requires an algorithm that looks deeply into the overall structure of relations among species, like a chess-playing program. Note that at each stage the algorithm of this paper is always comparing actual distances between species and choosing to cluster the two remaining species with minimum actual distance.

The "representation" of $V(S)$ as the total branch length of the reconstructed evolutionary tree of S comes directly out of the dynamic programming recursion (57). If $V(S \backslash g(S))$ is the total branch length of the hypothetical phylogenetic tree representing the evolution of $S \backslash g(S)$, then "adding" the species $g(S)$—the species which is closer to $h(S)$ than any other pair of species are to each other—"adds" incremental branch length $d(g(S), h(S))$. The geometric interpretation of the induction argument is completed by arbitrarily normalizing $V(S) \equiv d_0$ for $|S| = 1$. This amounts to normalizing the total branch length of the tree for $|S| \geq 2$ by making the (common) distance from the hypothesized outgroup species to the twig tip of any existing species of S equal to d_0, a given large constant.[21]

When equation (19) or (56) is applied to the distances from Table III, it yields the maximum likelihood evolutionary tree depicted in Figure I with the accompanying "as if" ultrametric distances of Table I. The tree is drawn in deterministic fashion by systematically utilizing, as seems reasonable, the information automatically generated at each stage of the algorithm about which of the two closest remaining species is the "representative" and which is the "link." The species "representing" the clade formed by each ancestor node is always placed farthest from the other clade from which the ancestor node has most recently diverged because there is a greater loss of diversity from subtracting the "representative" species than from subtracting the "link" species. Except for the first branch, where it is arbitrary which clade is placed to the left and which to the right, the convention naturally suggested by the algorithm completely determines species placement within the maximum likelihood tree and results in a com-

---

21. In the context of the maximum likelihood probabilistic model, the normalization convention essentially represents a harmless scaling away of a loose degree of freedom; survival probabilities are pinned down by assuming that the absolute probability that (any) one particular species "makes it" all the way through the hazards of the evolutionary maze to be alive today is an arbitrary, but irrelevant, constant $P_0 \equiv \exp(-d_0)$.

pletely deterministic two-dimensional topological representation of where existing species stand in the associated taxonomic tree.[22]

As a specific example, Figure I and Table I are derived from the primitive distance data of Table III in the following fashion. The two closest species of Table III are 5 and 6. When equation (56) is solved, it turns out that 5 is the "representative" and 6 is the "link." After 6 is eliminated, the two closest species (in the remaining five-by-five submatrix of Table III) are 1 and 2. Of this pair, (56), when it is solved, indicates that 1 is the "representative" and 2 is the "link." Eliminating 2, species 1 and 3 are the next closest, with 3 the "representative" and 1 the "link." The algorithm continues thus until only one species remains.

As reflected in Figure I, species 3 "represents" the "as if" clade {1,2,3}, and therefore is placed farthest from 4 in the "as if" clade {1,2,3,4}. Species 1 "represents" the "as if" clade {1,2} and therefore is placed farthest from 3 within the "as if" clade {1,2,3}. The "as if" ultrametric distances of Table I are constructed out of the "link" distances generated in making Figure I, and the value of the diversity function is the total branch length of the tree (up to an arbitrary normalization constant).

Although the underlying objects, whose diversity we seek, might not have been generated by any real evolutionary process, the algorithm nevertheless wants us to classify them *as if* they were the end products of a well-defined evolutionary phylogeny. The algorithm is always trying to tell an evolutionary story of diversity even though the diversity might not actually have been caused by genuine evolution.

It is important to be clear about the difference between the simulated evolutionary tree of the general case, whose artificial "as if" distances within the tree are ultrametric, and the genuine evolutionary tree associated with the true ultrametric distances of a true evolutionary process. When true underlying distances are ultrametric, any reasonable clustering method will reconstitute the genuine tree. In this case the simulated evolutionary tree and the genuine evolutionary tree are the same. But in the general case when actual distances are not ultrametric, the artificial evolutionary tree has some very different—and much weaker—properties from those possessed by a real evolutionary tree.

A genuine evolutionary tree is stable in the sense that the

22. Assuming, as seems reasonable, that multifurcations or exactly ultrametric distances are events of probability measure zero.

same tree skeleton remains no matter what species are subtracted in whatever order. The loss of diversity from a species becoming extinct can be myopically viewed at any stage as the length of the evolutionary branch of the species at that stage. When a species becomes extinct, the phylogenetic tree is diminished by snapping off the evolutionary branch of that species. But this analogy does not hold for an artificial evolutionary tree constructed in the general case of non-ultrametric underlying distances. This tree is just an approximation, albeit the "most likely" approximation. In the general case, there is only *one* species that satisfies the branch-snapping analogy. (In Figure I it would be species 6.) If any other species were eliminated (species 1 through 5 in Figure I), the loss of diversity is (generally) *not* the length of its branch. Even more problematical, the entire topology of the remaining tree may conceivably change if the artificial tree rearranges itself when a species is eliminated. In general, the diversity of a subset of species cannot be directly inferred from the artificial tree of the full set.

We can perhaps summarize this discussion by saying that the structure of the artificially induced evolutionary tree of the general case may be *fragile* (to elimination of species) if distances are far from being ultrametric, while the structure of the genuine evolutionary tree that arises from a real evolutionary process is *robust* (to elimination of any species).

VI. NICE PROPERTIES OF THE DIVERSITY FUNCTION

The rest of the paper essentially consists of a list of desirable properties that the proposed diversity measure (19) possesses. Not every item on the list is equally important; some are much more fundamental than others. All of the items seem reasonable, and it is difficult for me to think of any obviously desirable (yet attainable) properties of a diversity measure that have not been included.

As was pointed out previously, in the end the real argument in favor of the diversity measure (19) being proposed here is that it possesses many "nice" properties, while other candidates for a value of diversity function are "not nice" in one or more essential aspects. In that sense, the list of nice properties is critical to establishing the validity of $V(S)$. Again, this is a point that can only be appreciated by actually trying out other candidates for a value of diversity function and seeing where they fail to behave nicely.

The list of nice properties can be divided crudely into categories corresponding roughly to mathematical, taxonomic, ecological,

and economic aspects of the desired measure. After each item is introduced, a precise formulation of it is given; a proof is outlined or provided; and some attempt is made to motivate or apply the property.

The remainder of this section deals with properties categorized loosely as "mathematical."

1. MONOTONICITY IN SPECIES. Already covered in Section IV.

2. LINK PROPERTY. Already covered in Section IV.

3. TWIN PROPERTY. Suppose that some species k outside of S is identical to some species j belonging to S, meaning for some $j \in S, k \notin S$ that

$$d(j,k) = 0,$$

$$d(j,i) = d(k,i), \ \forall i \in S.$$

Then if k is added to S, there is no change in diversity:

$$V(S \cup k) = V(S).$$

Proof. Follows from the basic representation theorem.|

The (identical) twin property is an important statement about continuity of diversity in species addition or subtraction. If a species is added that is very closely related to an existing species, it should only raise the value of diversity by a very small amount that goes to zero in the limit as the added species becomes an identical twin with an existing species. The twin property should hold for any reasonable diversity function. If it does not hold, some obviously wrong allocation choices can be made because resources may be spent inappropriately on preserving species that have identical counterparts in no danger of extinction.

4. CONTINUITY IN DISTANCES. Let $|S| = |S'|$. Let $\phi(\cdot)$ be a one-to-one function mapping S onto S'.

Then, $\forall \epsilon > 0, \exists \delta > 0$ such that $\Sigma\Sigma |d(i,j) - d(\phi(i),\phi(j))| < \delta$ implies that

$$|V(S) - V(S')| < \epsilon.$$

Proof. Follows directly from the fact that $V(S)$ is the maximand of a linear programming problem (22), (23), whose parameters are continuous in distances.|

That the value of diversity function must be continuous in the underlying distances is an obviously necessary property. Surprisingly, this condition is sufficiently strong to eliminate some otherwise not completely unreasonable candidates for a diversity measure.

5. MONOTONICITY IN DISTANCES. Let $|S| = |S'| \geq 2$. Let $\phi(\cdot)$ be a one-to-one function mapping S onto S'.

Suppose that

$$d(\phi(i),\phi(j)) \geq d(i,j), \; \forall i \in S, \forall j \in S, i \neq j.$$

Then

$$V(S') \geq V(S).$$

Proof. Straightforward from the definition of $V(S)$.|

The above monotonicity property has an obvious interpretation that any meaningful diversity function based on pairwise distances should possess.

6. MAXIMUM DIVERSITY THAT CAN BE ADDED BY A SPECIES. If species k is added to collection S, then

(58) $$V(S \cup k) \leq V(S) + D(k,S), \; \forall S, \forall k \notin S,$$

where $D(k,S)$ is the *maximum* distance from point k to set S,

$$D(k,S) \equiv \max_{i \in S} d(k,i).$$

Proof. By induction, suppose that (58) holds $\forall S, |S| = n$. (It is certainly true for $n = 1$.) Now let S be any set with $|S| = n + 1$, and let k be any point $k \notin S$. From (19) applied to $S \cup k$, it follows that

(59) $$V(S \cup k) = \max \{d(k,S) + V(S),$$
$$\max_{i \in S} \{d(i,S \cup k \setminus i) + V(S \cup k \setminus i)\}\}.$$

From the induction assumption (58) applied to $S \setminus i$,

(60) $$V((S \setminus i) \cup k) \leq V(S \setminus i) + D(k,S \setminus i).$$

From first principles,

(61) $$D(k,S \setminus i) \leq D(k,S),$$

(62) $$d(i,S \cup k \setminus i) \leq d(i,S \setminus i).$$

Putting inequalities (60), (61), and (62) into (59) produces

(63) $V(S \cup k) \leq \max \{d(k,S) + V(S),$
         $\max_{i \in S} \{d(i,S\backslash i) + V(S\backslash i) + D(k,S)\}\}.$

Making use of (19), condition (63) reduces to

$$V(S \cup k) \leq V(S) + D(k,S),$$

which is the induction step to be proved. |

Inequality (58) represents an upper bound on the extra diversity that can be added by including a new species. The increment to diversity should be no more than the maximum distance from the new species to any old species in the set. Condition (58) is an intuitively reassuring result. Some reflection reveals that it cannot in general be strengthened to a tighter bound without putting more structure on the problem.

VII. TAXONOMIC ASPECTS

The next two properties are inspired by basic principles of taxonomy. Intuitively, there should be some fundamental connection between diversity and taxonomy, because both are ultimately based on related underlying concepts of similarity and difference. Indeed, this general duality is central to the paper. The following two specific items on the list of nice properties elaborate some basic features that it might be hoped a well-defined diversity measure would possess in the context of a meaningful taxonomy.

1. CLADE AGGREGATION. A *generalized clade* of species, denoted C, is defined to be a subset of S

$$C \subset S, \ |S| > |C| \geq 2,$$

that satisfies

(64) $d(i,S\backslash i) = d(i,C\backslash i), \qquad \forall i \in C,$

(65) $d(i,k) = d_k, \qquad \forall i \in C, \forall k \in S\backslash C.$

Let c stand for any species representing the generalized clade, meaning,

(66) $d(c,k) \equiv d_k, \qquad \forall k \in S\backslash C.$

Then

$$(67) \qquad V(S) = V(C) + V((S\backslash C)\cup c) - d_0.$$

Proof. By induction on the size of S, suppose that the clade aggregation property is true for $|S| = n$. (It is readily confirmed that clade aggregation holds for $n = 3$.) Now let $|S| = n + 1$. For convenience, normalize $d_0 \equiv 0$.

Decomposing (19) into two separate parts,

$$(68) \qquad V(S) = \max\{\alpha,\beta\},$$

where

$$(69) \qquad \alpha \equiv \max_{i\in C} \{d(i,S\backslash i) + V(S\backslash i)\},$$

$$(70) \qquad \beta \equiv \max_{k\in S\backslash C} \{d(k,S\backslash k) + V(S\backslash k)\}.$$

Using (64), equation (69) can be rewritten as

$$(71) \qquad \alpha = \max_{i\in C} \{d(i,C\backslash i) + V(S\backslash i)\}.$$

Using (65), (66), equation (70) can be rewritten as

$$(72) \qquad \beta = \max_{k\in S\backslash C} \{d(k,((S\backslash C)\cup c)\backslash k) + V(S\backslash k)\}.$$

Applying the induction assumption (67) to $S\backslash i$ with $i \in C$, we have

$$
\begin{aligned}
(73) \qquad V(S\backslash i) &= V(C\backslash i) + V(((S\backslash i)\backslash(C\backslash i))\cup c) \\
&= V(C\backslash i) + V((S\backslash C)\cup c).
\end{aligned}
$$

Plugging (73) into (71) yields

$$\alpha = V((S\backslash C)\cup c) + \max_{i\in C} \{d(i,C\backslash i) + V(C\backslash i)\},$$

which by (19) is equivalent to

$$(74) \qquad \alpha = V(C) + V((S\backslash C)\cup c).$$

Applying the induction assumption (67) to $S\backslash k$ for $k \in S\backslash C$, we have

$$
\begin{aligned}
(75) \qquad V(S\backslash k) &= V(C) + V(((S\backslash k)\backslash C)\cup c) \\
&= V(C) + V(((S\backslash C)\cup c)\backslash k).
\end{aligned}
$$

Plugging (75) into (72) yields

$$(76) \qquad \beta = V(C) + \max_{k\in S\backslash C} \{d(k,((S\backslash C)\cup c)\backslash k) + V(((S\backslash C)\cup c)\backslash k)\}.$$

Using first principles on (76) obtains

(77) $\beta \leq V(C) + \underset{k \in (S \backslash C) \cup c}{\text{maximum}} \{ d(k,((S \backslash C) \cup c) \backslash k) + V(((S \backslash C) \cup c) \backslash k) \}.$

Applying (19) to its right-hand side, expression (77) is equivalent to

(78) $\beta \leq V(C) + V((S \backslash C) \cup c).$

Combining (74) and (78) with (68), we have, at last,

$$V(S) = V(C) + V((S \backslash C) \cup c),$$

the induction step to be proved. |

A generalized clade, hereafter g.clade, is a subset of species, each of whose closest relative belongs to the g.clade, and each of whose distance from any species outside the g.clade is identical. A g.clade generalizes the concept of a clade to the case of non-ultrametric distances. The clade aggregation property decomposes the diversity of a set of species into two components: the diversity of the g.clade itself plus the diversity of all the other species aggregated with a single hypothetical representative of the g.clade. This result can be repeated for any number of g.clade subsets. The decomposability of g.clade subsets has significant computational and economic consequences. If preservation issues happen to concern only members of a particular g.clade, the analysis of resource allocation alternatives can safely proceed as if the g.clade were the fundamental universe of species, in isolation from non-g.clade members. To the extent that a set of species can be subdivided into g.clades, the computation of overall diversity is made much easier by taking advantage of clade aggregation.

2. ULTRAMETRIC DISTANCES REDUCE DIVERSITY THEORY TO PERFECT TAXONOMY THEORY. Already covered in Sections III and IV.

VIII. ECOLOGICAL DIVERSITY

This section applies the diversity function of the present paper to a highly stylized ecological community. Suppose that the underlying resources of the habitat support a total population of N units of biomass. Let there be l different "ecological-species." (In this section "ecological-species" means *real* biological species in the traditional sense.) The collection of individuals of ecological-species j will be denoted S_j ($j = 1, 2, \ldots, l$). It is assumed, symmetrically, that individuals of different ecological-species are equally diverse from each other, while individuals belonging to the same

ecological-species are identical. Let ecological-species j ($j = 1, 2, \ldots, l$) consist of N_j individuals, measured in common units of biomass. That is,

$$|S_j| = N_j.$$

If S is the collection of all individuals, then

$$|S| = N,$$

(79) $$S = \cup S_j,$$

$$N = \Sigma \, N_j.$$

Suppose that the habitat is modeled as a line segment of length N, which conceptually might represent a one-dimensional gradient of some ecological significance.[23] Ecological-species j occupies an ecological-species niche represented by a line segment of length N_j belonging to $[0, N]$. Each individual of the ecological-species occupies a single niche one unit long within its ecological-species niche. Figure II depicts the model representation for three ecological-species.

Think of N and all the $\{N_j\}$ as being indefinitely large and satisfying

(80) $$\lim_{N \to \infty} \frac{N_j}{N} = P_j > 0, \qquad \forall j = 1, 2, \ldots, l,$$

where

(81) $$\Sigma \, P_j = 1.$$

Conceptually, the $\{P_j\}$ are constants representing the relative biomass of each ecological-species.

Note that the numbers N and each of the $\{N_j\}$ can automatically be made indefinitely large in the requisite limiting way simply by making the biomass weight units and habitat distance units sufficiently small.[24] Actually, this is perhaps the best way to think of the limiting process (80) because it allows the total biomass and habitat to remain constant while rendering the density of all

---

23. For example, if the habitat is a salt marsh, the one-dimensional gradient might represent the degree of salinity. The dominant ecological species in a salt marsh are the two *Spartina* grasses. The ecological-species *Spartina alterniflora* is more tolerant of tidal immersion than *S. patens,* and hence there is a natural ordering of their niche location positions within the salinity gradient.

24. Think of metric tons of biomass becoming 10^3 kilograms of biomass, becoming 10^6 grams of biomass, becoming 10^9 milligrams of biomass, and so on.

N₁ = Niche of 1   N₂ = Niche of 2   N₃ = Niche of 3
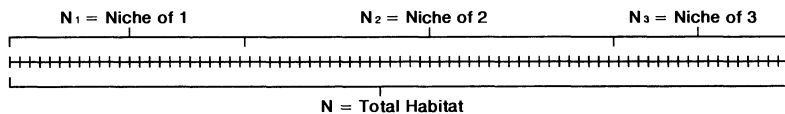
N = Total Habitat

FIGURE II
Ecological Species Niches

individuals to be uniform throughout their constant habitat-niche range.

Now "pretend" temporarily that each individual is like a separate species, whose ecological distance from any other individual is represented by how far apart they are on the niche line.

Let

$$\tilde{V}(S;N)$$

represent the naive "inflated" diversity function of S obtained from pretending that all individuals of S constitute different species.

Let

$$\tilde{V}(S_j;N)$$

represent the "false" diversity function of set S_j (in isolation) obtained from pretending that all individuals of S_j constitute different species. This "diversity" is completely spurious because it should be zero, since the set S_j consists of identical individuals.

It is natural to define the "true" diversity V^* of the model ecological community (per unit biomass) to be

$$(82) \qquad V^* \equiv \lim_{N\to\infty} (1/N)[\tilde{V}(S;N) - \sum_j \tilde{V}(S_j; N)]$$

subject to (80) holding in the limit.

Definition (82) corrects the double counting inherent in the "inflated" diversity of S by subtracting off the sum of "false" diversities of the l ecological-species sets. (In reality, the ecological-species sets should each have zero diversity.) The true diversity of S is the residual that remains after removing from the inflated diversity function of S all spurious diversity contributed by the ecological-species sets.

The widely used Shannon diversity index is

$$(83) \qquad\qquad H \equiv -\Sigma\, P_j \log_2 P_j.$$

THEOREM. Under the assumptions of the model,

(84) $V^* = H/2.$

Proof. We begin with the following.

LEMMA. Suppose that there is a collection A of M different species-individuals spaced one unit apart from each other along a line segment of length $M - 1$. (Set A is depicted in Figure III for the case $M = 5$.) Suppose that

$$M = 2^K + 1,$$

for some positive integer K.   Then

(85) $V(A) = (M - 1)/2 \log_2(M - 1) + (M - 1).$

Proof of Lemma. By induction on K, making repeated use of the basic representation theorem. The flavor of the proof can be given for $M = 5$, illustrated in Figure III. The distance between each pair of neighboring points is one. It is not difficult to show that of the nearest neighbors 1 and 2, the point cast out is 2, adding distance one. Similarly, between nearest neighbors 3 and 4, the point to be cast out is 4, adding distance one. These two operations "core out" points 2 and 4, leaving the set (1,3,5). Of the remaining nearest neighbors 1 and 3, the point to be "cored out" is 3, adding distance two and leaving end points 1 and 5, whose distance is four. Thus, $V(A) = ((1 + 1) + 2) + 4 = 8$, as given by (85). In the general case, the first coring operation "cores out" $(M - 1)/2$ points, adding total distance $((M - 1)/2) \cdot 1 = (M - 1)/2$ and leaving each of the remaining $((M - 1)/2) + 1$ points separated by a distance of two. The next coring operation "cores out" $(M - 1)/4$ points, adding total distance $((M - 1)/4) \cdot 2 = (M - 1)/2$ and leaving each of the remaining $((M - 1)/4) + 1$ points separated by a distance of four. Each subsequent "coring out" operation adds total distance $(M - 1)/2$. There are $\log_2(M - 1)$ such coring operations that it is possible to perform. Finally, one is left with the end points at distance $M - 1$ apart. Summing up all such distances yields (85). |
The remainder of the proof proceeds as follows. For all $j = 1, 2, \ldots,$ l, define

$$K_j(N_j) \equiv \log_2(N_j - 1).$$
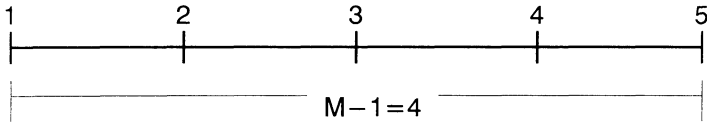
Define

$$K(N) \equiv \log_2(N - 1).$$

FIGURE III
Linear Equidistant Species

In the limit as N becomes indefinitely large while (80) holds, $K(N)$ and the $\{K_j(N_j)\}$ can be increasingly accurately approximated by integer values.[25]

Therefore, applying the lemma to (82),

$$(86) \quad V^* = \lim_{N \to \infty} \frac{1}{N} \left[\frac{N-1}{2} \log_2(N-1) + N - 1 \right.$$

$$\left. - \sum_j \left(\frac{N_j - 1}{2} \log_2 (N_j - 1) + N_j - 1 \right) \right].$$

Using (79), expression (86) can be rewritten as

$$(87) \quad V^* = \lim_{N \to \infty} \frac{1}{2} \left[\frac{2(l-1)}{N} + \frac{N-1}{N} \log_2(N-1) \right.$$

$$\left. - \sum_j \frac{N_j - 1}{N} \log_2(N_j - 1) \right].$$

Taking the appropriate limits and making use of (80), (81), expression (87) becomes

$$V^* = -\tfrac{1}{2} [\Sigma P_j \log_2 P_j],$$

the result to be proved.|

Theorem (84) ties together two concepts of diversity. The Shannon-Weiner formula (83) is a famous measure of information or of entropy. It has been widely used in ecology as an index of diversity for ecological communities.[26]

The concept of diversity being used in this paper is in principle more general than ecological diversity, which would appear as if it

25. I present here what is sometimes called a "physicist's proof." A fully rigorous proof that spells out in detail all the fine points of the limiting argument constitutes several pages that would seem disproportionately out of place here. The reader who cares should be able to fill in the details, as they are straightforward.

26. See Pielou [1977] for an outstandingly lucid treatment of ecological diversity issues, including an axiomatic derivation of the Shannon diversity index. Magurran [1988] contains a comprehensive survey of ecological diversity indexes.

ought to be some kind of special case. It is therefore a gratifying and reassuring result that, for at least one reasonable interpretation of an ecological community, the diversity function of this paper reduces to the well-known Shannon index.

IX. Economic Considerations

What follows next is a list of "economic items." The basic philosophy behind this approach is simple. Imagine the following thought experiment. Suppose, without having a formal definition of diversity, that we know intuitively the solution or the solution form of some particular, if hypothetical, diversity-maximizing resource-constrained allocation problem. Then a formally defined diversity function should yield the intuitively valid solution or solution form. In this way, intuitive thought experiments about resource allocation problems involving diversity may be used to validate the appropriateness of particular diversity functions.

1. FAVOR THE MORE DISTANTLY RELATED SPECIES. Suppose that $1 \in S, 2 \in S$, and

$$d(1,i) < d(2,i), \quad \forall i \in S \backslash 1 \backslash 2.$$

Then

$$V(S \backslash 2) < V(S \backslash 1).$$

Proof. By induction on $n = |S|$. Details left as an exercise.|

If one species is unambiguously more distantly related to the rest of the population than another, it should always be preserved over the other, ceteris paribus.

The above result expresses one sense in which the more distantly related species should be favored. Another sense is the following.

Let $|S| = n$ (≥ 2). Let B be the set of all $n(n-1)/2$ possible pairs of different species $(i,j) \in S, i \neq j$. Let $(i^*, j^*) \in S$ solve

$$d(i^*, j^*) = \underset{i,j \in S}{\text{maximum}} \, d(i,j).$$

Then

$$V(i^*, j^*) = \underset{(i,j) \in B}{\text{maximum}} \, V(i,j).$$

Proof. Follows directly from the relevant definitions.|

If only *two* species can be preserved out of a set, they should be the most distantly related pair.

2. IRRELEVANCE OF EQUALLY DISTANT RELATIVES. Suppose that there is a species $k \notin S$ satisfying

(88) $$d(k,i) = d, \qquad \forall i \in S.$$

Then

(89) $$V(Q \cup k) = V(Q) + d, \qquad \forall Q \subseteq S.$$

Proof. Condition (88) implies that

(90) $$d(k,Q) = D(k,Q) = d, \qquad \forall Q \subseteq S.$$

Conclusion (89) then follows from combining (21) and (58) with (90).|

If one species is equally related to all other species under consideration, its presence or absence is irrelevant to decisions concerning the preservation or extinction of any subsets of the other species.

3. RULE OF THE SNAKE. Let $S = \{1,2,\ldots,n\}$. Suppose that there exists a strict ordering,

$$d(i,k) \leq d(j,k) \qquad \forall i,j,k \in S, \qquad i < j.$$

Then

$$V(S) = \sum_{i=1}^{n} d(i-1,i).$$

Proof. By induction on n. Left as an exercise.|

Suppose that there is an unambiguous sense in which species can be ranked from more closely to more distantly related (to the rest of the group). Call the more distantly related species more "valuable." Then the total value of diversity of the group should be the sum of pairwise distances along a chain connecting the most valuable species to the next most valuable, to the next most valuable after that, and so forth.

The basic intuition should be clear. A concept like "the most valuable species" is generally problematical because it is inherently a global property; the most valuable species can change as other species in the set become extinct. However, in this particular case the concept makes sense. The most valuable species is the farthest distant from the others—by any reckoning. When the most valuable species has been eliminated, value has been lost equal to the distance from the most valuable to the next most valuable

species—and so forth along a chain-linked snake whose head is the most valuable species and tail is the least valuable.

4. ADDITIVITY PROPERTIES OF INDUCED UTILITY FUNCTIONS. Let $k \notin S$ be a hypothetical "reference" or "comparison" species for S. Define

$$U(i) \equiv d(k,i), \qquad \forall i \in S,$$
$$U(i,j) \equiv V(i,j,k), \quad \forall i,j \in S,$$
$$U(S) \equiv V(S \cup k).$$

With additive separability, total utility is the sum of individual utilities:
Suppose that

$$U(i,j) = U(i) + U(j), \qquad \forall i,j \in S.$$

Then

$$U(S) = \sum_{i \in S} U(i).$$

With substitutes, total utility is less than the sum of individual utilities:
Suppose that

$$U(i,j) \leq U(i) + U(j), \qquad \forall i,j \in S,$$
$$U(i',j') < U(i') + U(j'), \quad \text{for at least one pair } i',j' \in S.$$

Then

$$U(S) < \sum_{i \in S} U(i).$$

With complements, total utility is more than the sum of individual utilities:
Suppose that

$$U(i,j) \geq U(i) + U(j), \qquad \forall i,j \in S,$$
$$U(i'',j'') > U(i'') + U(j''), \quad \text{for at least one pair } i'',j'' \in S.$$

Then

$$U(S) > \sum_{i \in S} U(i).$$

Proof. By induction on $n = |S|$. Left as a nontrivial exercise. (Hint: use the fact that the diversity of a triangle is the sum of its shortest and longest sides, and apply the fundamental representation theorem to $S \cup k$.)|

The intuitive explanation of the above family of results is as follows.

Suppose that we are interested in preserving certain collections of objects, each individual member of which may possess some inherent value. An example might be architectural landmark buildings. Let k stand for a reference object, not of interest for its own sake, against which the inherent utility of each of the other objects is calibrated. In the example, k might stand for an ordinary, average, nonlandmark building. The inherent utility of k alone is normalized to be zero; the utility of any object of interest is always measured relative to the reference k and is made equal to the "distance" of that object from k. In the example, the stand-alone value of any particular landmark building is determined relative to the ordinary building k, and is identified with the imputed distance of that landmark building from k.

Now if two objects share very little in common, say because they represent entirely different periods or styles, their joint utility should be approximately the sum of their individual utilities. On the other hand, if two buildings substitute for each other, say because they are both examples of a particular style, the utility of the pair should be less than the sum of their individual utilities, even though one building might have higher individual utility because it is a better preserved specimen than the other. Conversely, if two objects complement each other, like a matched pair of gloves, their joint value should exceed the sum of their individual values.

A good theory of diversity should automatically pick up these themes and integrate them into its main body. The framework of this section can be viewed as an extension of the theory of diversity to the case where objects have an intrinsic value or utility. If they are to be trustworthy evaluators, these induced utility-diversity functions should possess basic properties of utility functions. When individual utilities are additively separable, total utility should equal the sum of individual utilities. With diminishing returns to combining species, total utility should be less than the sum of individual utilities. And when there are increasing returns to combining species, total utility should exceed the sum of individual utilities.[27]

27. This section represents a tentative first step in the direction of trying to show how the approaches to diversity used in economic models might be viewed as special cases within the framework of the current paper. A full development of this theme seems a worthy subject of future research.

5. MIN-LOSS EXTINCTION. Without loss of generality suppose that 1 $\in S$ and $2 \in S$ and

$$d(1,2) \leq d(i,j) \ \text{ for } \forall i,j \in S, \ i \neq j.$$

Then either $i' = 1$ or $i' = 2$ is the solution of

$$V(S \backslash i') = \max_{i \in S} V(S \backslash i).$$

Proof. Min-loss extinction is a direct consequence of the fundamental representation theorem. |

Suppose that S contains n species, but the underlying "budget constraint" allows exactly $n - 1$ of these species to survive with certainty, while exactly one species must perish with certainty. If there is a choice, which one of the n species should be allowed to go extinct so that the other $n - 1$ might be preserved? The intuitive answer is that one of the two most closely related species in the set should be the single species that is extinguished.

The min-loss extinction property is a strong general criterion for discriminating among proposed diversity functions.

At the end of Section III a powerful extension of this concept was stated for the special case of ultrametric distances.

X. CONCLUDING REMARKS

This paper is difficult to summarize. I hope it demonstrates that, by starting with the basic concept of a diversity function, a powerful, unified, general theory of diversity can be developed, which yields many useful insights over a wide range of applications.

HARVARD UNIVERSITY

REFERENCES

Caccone, A., and J. R. Powell, "DNA Divergence Among Hominoids," *Evolution, 43* (1989), 925–42.
Erwin, T. L., "An Evolutionary Basis for Conservation Strategies," *Science, 253* (1991), 750–52.
Felsenstein, J., "The Number of Evolutionary Trees," *Systematic Zoology, 27* (1978), 27–33.
——, "Phylogenies from Molecular Sequences: Inference and Reliability," *Annual Review of Genetics, 22* (1988), 521–65.
Gondran, M., and M. Minoux, *Graphs and Algorithms* (New York: Wiley, 1984).
Jardine, N., and R. Sibson, *Mathematical Taxonomy* (New York: Wiley, 1971).
Krajewski, C., "Conservation Phylogenetics: A Focal Role for Molecular Systematics in the Biodiversity Crisis," Southern Illinois University, mimeo, 1991.
Li, W. H., and D. Graver, *Fundamentals of Molecular Biology* (Sunderland, MA: Sinauer, 1991).

Magurran, A. E., *Ecological Diversity and Its Measurement* (Princeton, NJ: Princeton University Press, 1988).

May, R., "Taxonomy as Destiny," *Nature, 347* (1990), 129–30.

Nei, M., *Molecular Evolutionary Genetics* (New York: Columbia University Press, 1987).

Oldfield, M. L., *The Value of Conserving Genetic Resources* (Washington, DC: U. S. Department of the Interior, 1984).

Pielou, E. C., *Mathematical Ecology* (New York: Wiley, 1977).

Sneath, P. H. A., and R. R. Sokol, *Numerical Taxonomy* (San Francisco: W. H. Freeman, 1973).

Solow, A. R., J. M. Broadus, and S. Polasky, "On the Measurement of Biological Diversity," Woods Hole Oceanographic Institution mimeo, 1991, in *Journal of Environmental Economics and Management,* forthcoming.

Swofford, D. L., and G. J. Olson, "Phylogeny Reconstruction," in D. M. Hillis and C. Moritz, eds., *Molecular Systematics* (Sunderland, MA: Sinauer, 1990).

Vane-Wright, R. I., C. J. Humphries, and P. H. Williams, "What to Protect? Systematics and the Agony of Choice," *Biological Conservation, 55* (1991), 235–54.

Weitzman, M. L., "A Reduced Form Approach to Maximum Likelihood Estimation of Evolutionary Trees," HIER working paper, 1991a.

―――, "What to Preserve? An Application of Diversity Theory to Crane Conservation," HIER working paper, 1991b.