
Diversity functions

Martin L. Weitzman

1.1 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 need a more-or-less consistent and usable measure of the value of diversity that can tell us how to trade off one form of diversity against another.

It would be naive to expect that resolution of real-world conservation choices will reduce to some mechanical application of diversity functions. Yet, I would argue, it is still useful to think in terms of a model that might serve as a paradigm for guiding and informing conservation decisions, even if the model must be at a high level of abstraction. When diversity cannot be defined even under ideal circumstances, the concept itself is suspicious. For this reason alone, it behooves us to specify a diversity function at least for some “ideal” case.

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 ac-

tions that can be taken, represented by budget constraints or other limitations. Each feasible action induces a probability distribution for what survives, and for how long (there is, perhaps, a significant amount of correlation involved). The optimal conservation 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 components). This is in the form of a classical constrained optimisation 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.

1.2 The nature of the problem

There is an immediate issue of defining the proper unit of analysis for the collection whose diversity is to be determined. It is not transparently clear in all conservation settings at what level the diversity problem should be attacked. In principle, diversity could be measured at the individual level, the species level, the community level, the ecosystem level, or even some other levels. (The Nature Conservancy usually takes its "mapping units" or "elements" to be species or communities.) I do not have a good resolution to the problem of which level is the most appropriate for performing diversity analysis. In principle, any level might be chosen so long as the methodology is consistently followed at that level.

The abstract form of the general problem can now be stated. The "elements" are basic units that it is desired be preserved in the name of diversity. (There could also be some direct net benefits from some elements.) There is some notion of the joint probabilities of extinction of the various elements if no action is taken. Next, there are preservation "actions" that can influence the various probabilities of extinction at some cost. For example, projects might consist of buying up and preserving various specific sites. Some "diversity function" evaluates the diversity of any given deterministic collection of elements. Conceptually, the diversity function, on which this paper concentrates, is the most difficult part of the problem. If a diversity function is well defined, and probabilities of extinction are known, an expected diversity function can be defined. An expected diversity function is basically the sum of the deterministic diversity function of various collections of species weighted by the existence probabilities of the various collections. The basic aim might be taken to maximise expected present discounted diversity (plus any net direct values of the elements), subject to conservation budget constraints. The set-up is analo-

gous to a capital budgeting problem, except that the objective contains expected diversity.

For convenience and consistency, in what follows I will largely employ biological metaphors. However, 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 it could stand for an individual, a subspecies, a specimen, an object, a community, or almost anything else – depending on the context.

Actually, one of the most useful interpretations is that a "species" corresponds to a "library." Conceptualising the basic problem in terms of preservation priorities among "libraries" is useful for at least two reasons. First of all, a library, at least in the abstract, tends to be a more or less neutral object that does not conjure up such strong emotional images as some other metaphors. Second, and more significantly, concentrating on libraries and the books they contain can help us to focus more sharply on the issue of what should be meant by a diversity function.

In what follows, then, the word "species" may be interpreted as some generalisation of the word "library." The reader who wants a specific image may find it useful to think of a species as standing for a library.

Suppose, then, there is some set S containing n member species (or n libraries). The basic question is how to measure the diversity of S . The appropriate diversity function will be denoted in this paper $V(S)$.

It is important to realise that there is unlikely to be a universally best definition of diversity. This is just common sense reasoning by analogy. There is not in statistics a universally best definition of central tendency, or of dispersion. Nor is there in economics a universally best definition of income inequality, or of welfare, or of industrial concentration. The appropriate definition depends upon the assumptions behind the specific intended application. However, just as in statistics or economics the field tends in practice to narrow down to only a few good candidates, so too I will try to argue, it is not so easy to find many good candidates for a diversity function. I will try to explain and justify my own formulation of a diversity function. Although I have not yet come across a formulation that satisfies me as much as the one I present here, the field is young and it is not to be excluded that some fresh approach might yield new insights.

What should one mean by a "good" definition of a diversity function? I think, as with the quantification of any concept, there are two general criteria.

First of all, the definition should be *a priori* sensible in that it embodies

an intuitively plausible formulation that does not immediately admit of seriously damaging counterexamples to the basic underlying idea. Obviously, this criterion contains a subjective element. Second, and perhaps more critically, there should be some special case, hopefully a *sensible* special case, forming the central paradigm, for which the particular formulation is exactly the right answer to a rigorously well-posed problem. The definition in the general case then becomes seen as an appropriate abstraction of the basic concept to a situation where the problem is less rigorously stated than it is in the central paradigm.

The approach described above is consistent with standard statistical methodology. Basically, there is a rigorous model, which works exactly for an idealised situation that is not excessively bizarre even though it may not precisely characterise the real world. Additionally, the model itself seems sensible on heuristic grounds for the general case. I cannot here test the robustness of the model itself, because that would require a more general meta-model with a more general meta-definition of diversity, which I do not have. The most that can be said at this stage is that my definition of diversity works exactly for a nonbizarre central case, it has some nice properties and makes heuristic sense in the general case, and it might be hoped to have some robustness properties if one knew how to formulate them properly.

I want to start with the rigorous model of a nonbizarre special case. That is, I want to lay out a particular model of an idealised situation where it is really fairly clear what we should mean by a diversity function. The particular model is what I will call the “bead model” of an evolutionary branching process.

1.3 The bead model of an evolutionary branching process

Consider a treelike branching process such as depicted in Figure 1.1. The “species” 1 through 6 are depicted as twig tips at the end of the tree. Species evolve by descent with modification via an evolutionary branching process, which is described as follows.

Any species consists of the same very large number M of tiny beads strung together on a string. If the primary interpretation of a “species” is a “library,” then the “beads” stand for “books.” A species is essentially identified with its string of beads, just as a library is identified with its collection of books. The beads are accumulated over time by being drawn from an infinitely large sample pot of different beads. At each unit of time, for each species existing at that time, exactly one bead is independently drawn from the infinitely large sample of different beads and attached to the head of the string. Simultaneously, exactly one bead is dropped from

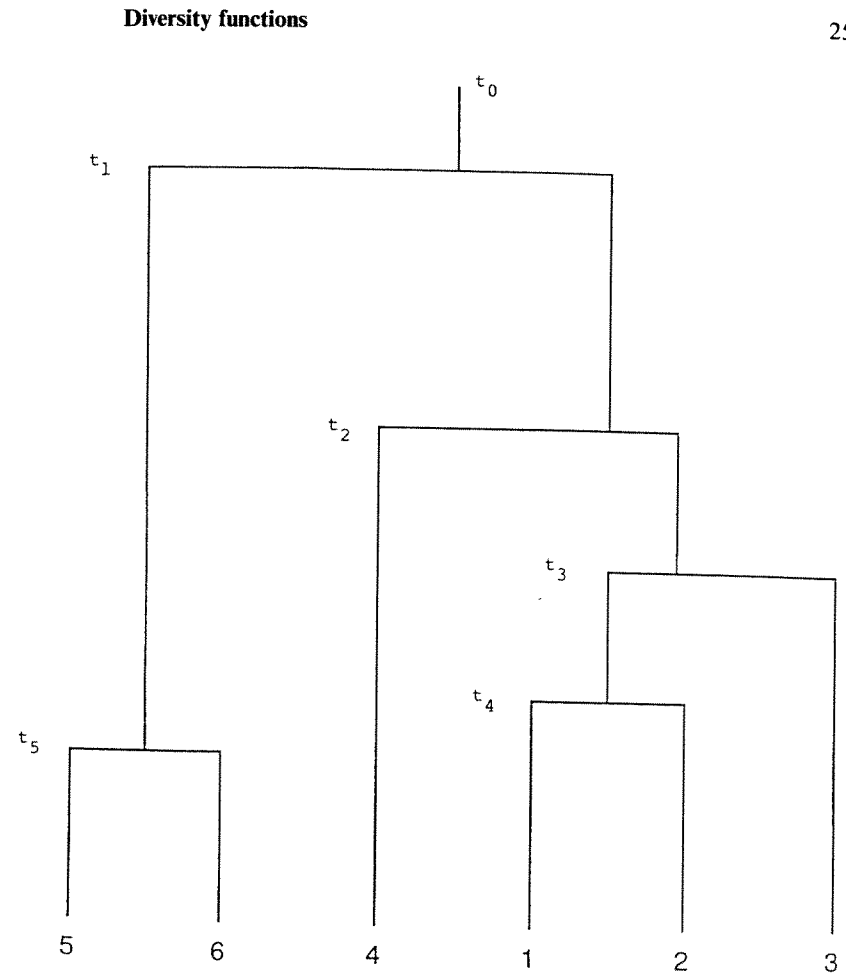


Figure 1.1. The maximum-likelihood tree representation.

the tail of the string. (Other descriptions are possible, but this one is the simplest.)

At time t_0 , just one prototype ancestor species exists. From time t_0 to time t_1 exactly $t_1 - t_0$ different new beads are accumulated at the head of the string, while the same number of old beads are discarded from the tail.

Then at time t_1 , a bifurcation into two ancestor species occurs. One of these is the common ancestor of $\{5, 6\}$. The other is the common ancestor of $\{1, 2, 3, 4\}$. Each of these two ancestor species, which are thought of as separate, now begins *independently* to accumulate different new beads at the head of the string, one per unit time, while simultaneously discarding one bead per unit time from the tail of the string. The next bifur-

cation occurs at time t_2 . At that time, the ancestor species of {5, 6} and of {1, 2, 3, 4} differ by exactly $t_2 - t_1$ beads, which they have independently acquired during the time duration $t_2 - t_1$.

At time t_2 , the common ancestor of {1, 2, 3, 4} bifurcates into two species. One of these is the common ancestor of {4}. The other species is the common ancestor of {1, 2, 3}. From time t_2 to t_3 , each of the three ancestor species {5, 6}, {4}, and {1, 2, 3} is *independently* accumulating different new beads at the head of their strings, one per unit time, while simultaneously discarding one bead per unit time from the tails of their strings. This phase ends at time t_3 , when ancestor species {1, 2, 3} bifurcates into ancestor species {1} and ancestor species {2, 3}. At that time t_3 , ancestor species {1, 2, 3} differs from ancestor species {4} by $t_3 - t_2$ beads, while ancestor species {1, 2, 3} and ancestor species {4} both differ from ancestor species {5, 6} by $t_3 - t_1$ beads.

The evolutionary branching process described above ends at the present time with the six currently existing species shown in Figure 1. 1. The last bifurcation that occurred was at time t_5 , when ancestor species {5, 6} split into ancestor species {5} and ancestor species {6}.

The model of an evolutionary branching process described above is of course an idealisation. Actual evolution differs in many important ways. Nevertheless, as an abstraction, the bead model captures the essential idea of descent with independent modification along reproductively isolated lineages. Supposing for the sake of argument that the model is a true description of how the species evolved, what does it tell us about the appropriate definition of diversity?

Each of the species 1 through 6 consists of a long string of beads of identical length. It seems natural to define the difference or distance between any two species as the number of beads that are different between them. With this definition, the distance between any two species can be read from the corresponding genealogical tree of Figure 1.1 as the time back to their nearest common ancestor. The number of beads by which two species differ is equal to the time elapsed from their most recent common ancestor because that is exactly the time period over which the different beads have been independently accumulated by the two species.

One possible definition of diversity in the present context is the total number of *different* beads contained in the collection. Some reflection should reveal that, in the present model, diversity under this definition equals the length of the associated taxonomic tree. By the length of an evolutionary tree, I mean the total lengths of all its vertical branches, including the branch of the common ancestor of the entire family back to some unspecified outgroup. The reader should confirm that the number of different beads represented by the six existing species of Figure 1.1 is

indeed equal to the total length of the associated tree under the bead model of an evolutionary branching process being assumed. Thus, under extreme simplifying assumptions about the nature of the underlying evolutionary process, if the diversity of a collection of species is defined to be the total number of different subunit-beads, then diversity equals the total branch length of the associated taxonomic tree.

There is an equivalent way of describing diversity in the above structure that is useful because part of it generalises. Without trying to be overly formal here, think heuristically of the operation of the bead process as a kind of "creation machine." Then the diversity of a collection is the number of "operations" required of the "creation machine" to make the collection. Later it will be shown how a more rigorous definition of the technology of a creation machine and its operations can be used to define diversity in the general case as the work required by the creation machine to make the diverse collection.

The preceding definition of diversity can be rephrased in terms of a "hierarchical search" procedure. In this context, think of a bead as a book. Then each species is like a library of M books. Any two libraries may contain certain books in common, and some that are different. Suppose every book embodied in the collection of species must periodically be searched to find, e.g., the appearance of a certain phrase. What is the best catalogue hierarchy for organising such periodic searches through all the books in the collection?

Some reflection reveals that the optimal hierarchical search structure is exactly the genealogical structure arising from the bead model that generates the evolutionary branching process. The highest level catalogue contains all books commonly held by every library {1, 2, 3, 4, 5, 6}. The next level of catalogue contains all books held in common by libraries {1, 2, 3, 4}, but not in {5} and {6}. The next catalogue after that contains books in {1, 2, 3} but not {4}. Then {1, 2} but not {3}, then {1} or {2} but not both, then {5} or {6} but not both.

The optimal way to search every book is to first search the books in the highest catalogue, then the next highest, then the next highest after that, and so forth, until every book is searched. This hierarchical search procedure takes minimal time among all the alternatives because it completely avoids the redundancy of searching the same book twice in two different libraries or catalogues. The total time required to examine all the books in an optimal search hierarchy is here exactly the total branch length of the associated genealogical tree (equal to the total number of different books). Thus, an alternative definition of diversity is the minimal time for a complete hierarchical search. This definition is useful because it can form the basis for generalising the concept of diversity to a situation

where the species, or libraries, do not conform to the bead model of an evolutionary branching process.

There is also a probabilistic interpretation, that can be motivated by a somewhat more operational criterion, but which yields the same essential identification of diversity with the total branch length of the associated taxonomic tree.

Suppose that we are looking for some desirable property, like a new source of food or medicine. Each collection of species can be viewed as a kind of natural portfolio of future options for the desirable property. Suppose that if the desirable property exists in a species, then it will be found in one or more of the beads out of which that species is composed. (Or, if the desirable property exists in a library, it will be found in at least one of the library's books.) Suppose that the probability of any particular bead having the desirable property is independent of any other bead having the desirable property and is equal to some small positive number ϵ . The probability that any bead does *not* contain the desirable property is then $\lambda = 1 - \epsilon$.

Now we can calculate the probability that the entire collection of species does *not* contain the desirable property. Let L be the number of different beads in the collection of species, equal, as we have seen, to the total branch length of the corresponding taxonomic tree. Then the probability that the entire collection does not contain the desirable property is $P = \lambda^L$. A not unreasonable definition of the diversity of a collection of species might be the negative logarithm of the probability that the collection does not contain the desirable property. By this definition, diversity is kL , where $k \equiv -\log\lambda$ is a positive constant. Thus, either concept of diversity yields essentially the same construct – namely diversity equals the total branch length of the associated taxonomic tree.

The above reasoning gives a powerful way of thinking about the loss of diversity that accompanies extinction events, at least for the bead model of an evolutionary branching process that has been presented.

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 subevolutionary 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 visualisation of the effects of various combinations of species losses on diversity in the special case of perfect taxonomy based on the bead model of an evolutionary branching process.

A simple example may help to illustrate the basic issues. In Figure 1.1 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 vanishes. 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 diversity as a conceptual aid for analysing policy options concerning preservation of diversity.

The previous reasoning can be pursued further to yield some not so obvious insights about conservation policy. Just to emphasise the abstract nature of the problem, suppose here we are talking about libraries. We have already mentioned how an expected diversity function can be defined when there is uncertainty. An expected diversity function is basically the sum of the deterministic diversity function of various collections of libraries, weighted by the existence probabilities of the various collections. Assume, as a simplification, that all survival probabilities are independent. Suppose the aim is to maximise expected diversity.

Consider the following numerical example. The numbers have been chosen to make the point sharply, but the point itself is quite general.

Referring again to Figure 1.1, let the survival probability of library 5 be $P_5 = .98$. Suppose the survival probability of library 6 is $P_6 = .02$. Library 5 might be called a relatively "safe" library, while library 6 is relatively "endangered."

Suppose now we consider the possibility of changing underlying resources to "shift" .01 of survival probability between libraries 5 and 6. Which of the following three alternatives yields the highest expected diversity?

Alternative	P_5	P_6
1. Status quo	.98	.02
2. Endangered library <i>more</i> endangered	.99	.01
3. Endangered library <i>less</i> endangered	.97	.03

I think it is fair to guess that most conservation-minded people would favour 3, the option that increases the survival probability of the endan-

gered library at the expense of the safe library. Actually, expected diversity is *minimised* by alternative 3, while it is *maximised* by alternative 2.

The reason for this counterintuitive result can be explained as follows. For simplicity I just compare situations 2 and 3. The probability that *both* libraries *survive* is $(.99)(.01) = .0099$ in situation 2, while it is $(.97)(.03) = .0291$ in situation 3. This would seem to turn the calculation in favour of 3 and is probably what accounts for the intuition that 3 yields higher expected diversity than 2.

However, the probability that *both* libraries become *extinct* is $(.01)(.99) = .0099$ in situation 2, while it is $(.03)(.97) = .0291$ in situation 3, which is an exact reversal of the previous calculation.

Now it would, of course, be good for diversity to have both libraries 5 and 6 survive. But it would be a significant disaster for diversity if both libraries 5 and 6 went extinct, because a whole lineage of unique books would then have been extinguished. Therefore, other things being equal, the analytical preservationist favours making the safe library safer at the expense of making the endangered library more endangered, because a whole line may therefore be made safer – if a one to one tradeoff of survival probabilities is possible. Although this example rests upon specific assumptions, I believe it offers some relevant insights into conservation policy that could not easily be made outside the diversity function framework.

The perfect taxonomy structure induced by the bead model of an evolutionary branching process allows yet other powerful insights into the form of an optimal conservation policy. Consider, for example, the following idealised situation involving sharply posed preservation issues in such a context. This might be called the “Noah’s ark problem.”

Let the set S consist of n species denoted by $i = 1, 2, \dots, n$. 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 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 the cost of preserving species i with probability x_i is equal to $c_i x_i$. (In the Noah’s ark interpretation, c_i is the room in the ark taken up by the pair of species i .) Let the row n -vector of cost coefficients be $C \equiv (c_i)$. Let the total preservation budget be A . (A is the size of the ark.)

The simplest form of a constrained expected diversity maximising problem might be formulated as:

$$\text{maximise } \phi(X)$$

subject to:

$$\begin{aligned} CX &\leq A \\ 0 &\leq x_i \leq 1 \quad \text{for } i = 1, 2, \dots, n \end{aligned}$$

The previous constrained optimisation problem is well-defined, but it looks like a combinatoric nightmare. Actually, in the case of the bead model a simple myopic algorithm is available for solving the problem. 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 at some iteration the subset $Q \subseteq S$ of species exists with probability one, while the subset $S \setminus Q$ of species is extinct or exists with probability zero. ($S \setminus Q$ stands for the set S minus the set Q .) Suppose that the budget constraint is not being met:

$$\sum_{i \in Q} c_i x_i > A$$

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

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

Distance $d(j, Q)$ from point j to set Q is understood in the usual sense to be the distance from j to the element of Q closest to j .

The probability x_j is then brought down continuously from one towards 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 .¹ 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 for the bead model of evolutionary branching such a myopic algorithm yields an optimal policy in the sense of satisfying the original optimisation problem. 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 sig-

¹ Note that this will change some of the remaining $\{d(i, Qi)\}$ coefficients.

nificance 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.

I hope I have been able to present a fairly convincing argument that for the special bead model of an evolutionary branching process there is a moderately compelling case for identifying the diversity of a collection of species with the length of the associated genealogical tree. Unfortunately, the bead model is an extreme abstraction of an idealised evolutionary process. It provides a useful construct within which it really is fairly clear what we should mean by a diversity function. The difficult question, to which I next turn, is what to call a diversity function for a situation where the bead model is not strictly applicable.

Remember that the "distance" $d(i,j)$ between libraries i and j is the number of books different between i and j . In the bead model, all distances are "ultrametric," meaning that for any three libraries i, j, k ,

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

Ultrametric distances have the enormously attractive property that they can be completely represented by a tree structure. Conversely, any (rooted, directed) tree defines a set of distances that are ultrametric.

This can be seen readily from Figure 1.1. The "distance" between any two species is represented in Figure 1.1 as the time back to the most recent common ancestor. Equivalently, this distance represents the collection of beads or books that are different between the two species or libraries.

I have tried to argue in this section that when distances are ultrametric, the leading candidate for a diversity function really should be fairly clear. With ultrametric distances, diversity is the total branch length of the associated tree. A variety of approaches or views support this interpretation. Furthermore, when distances are ultrametric a rather powerful theory can be developed to give insight into the nature of strategies that would maximise expected diversity. At this point we must address the issue of defining a diversity function in the more general case when distances are not ultrametric. We will proceed by attempting to generalise from the ultrametric case.

1.4 Diversity in the general case

Suppose we continue to think of a species as a collection of M beads on a string or M books in a library. Only now, the collection is not necessarily derived from the bead model of an evolutionary branching process, or

distances are not necessarily ultrametric. How are we then to define the diversity of a group of species?

As before, the distance between species i and j is the number of beads or books that are different between them. Here the distances $\{d(i,j)\}$ are taken as given data. As is traditional, the distance $d(j,Q)$ from the point j to the set Q is equal to the distance from j to the element of Q closest to j .

It is only rarely that distances are ultrametric. Far more frequently, distances are not exactly as if they are derived from the bead model. We are typically confronted initially with a situation where the given pairwise symmetric dissimilarity-distance measures are not ultrametric and are therefore not consistent with the bead model of an evolutionary branching process.

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

$$V(S) = \max_{i \in S} \{V(S \setminus i) + d(i, S \setminus i)\} \quad (1-1)$$

The dynamic programming Equation (1-1) is the centrepiece of the present approach to diversity. The solution of Equation (1-1) is unique once the initial conditions

$$V(i) \equiv d_0 \quad \forall i \quad (1-2)$$

are specified for any d_0 . Depending upon the particular application, it is typically most convenient to normalise 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 of Equation (1-1). These axiomatic treatments are suggestively motivating, as I hope to indicate. However, the real argument for the diversity measure being proposed here is that it "works" fairly well – 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 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

$$V(Q \cup j) \geq V(Q) + d(j, Q) \quad \forall Q \quad \forall j \notin Q \quad (1-3)$$

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

The monotonicity in species Condition (1-3) 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 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 neighbour 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 (1-3). There are at least two ways to add a supplementary condition that would make the Inequality (1-3) hold so "tight" that it yields, in effect, the dynamic programming Equation (1-1).

The first approach is the most direct. View Condition (1-3) as a potentially very large set of constraints that must hold for *all Q* and for *all j*. Impose the uniform initialising Condition (1-2). Then simply define the diversity of *S* to be the minimum possible $V(S)$ that satisfies Equations (1-3), (1-2).

The reason this direct approach yields the dynamic programming recursion (1-1) is as follows. Suppose, by induction, the diversity functions $\{V(S_i)\}$ have been defined for all *i* belonging to *S*. Then the smallest possible value for the diversity of *S* that would be consistent with Condition (1-3) must satisfy the condition:

$$V(S) \equiv \text{minimum } V \quad (1-4)$$

subject to:

$$V \geq V(S_i) + d(i, S_i) \quad \forall i \in S \quad (1-5)$$

It is straightforward to confirm that the solution of Equations (1-4), (1-5) is Equation (1-1), 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 Equation (1-3) 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 rooted directed branching representation of the evolutionary process. For each species, the length from root to twig tip in this branching process is the same number *M*. 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 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 simulta-

neously satisfy, for *all Q* and for *all j*, the basic consistency conditions of Equation (1-3).

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 Equations (1-4), (1-5), and, by extension, Equation (1-1). Thus, Equation (1-1) has the interpretation of describing the number of steps required to generate the most parsimonious "minimal evolution" branching structure that gives rise to the species of *S*.

This evolutionary metaphor can be recast as an insightful story about the cost of "making" diversity. Suppose the *n* objects each consist of *M* spaces or positions. Every position is filled with a particular colour, letter, flavour, codon, symbol, or whatever, depending on the context. Each object is in effect a mosaic of symbols. The "distance" $d(i, j)$ between objects *i* and *j* is the number of positions of *i* and *j* which have different symbols in them.

Think of a symbol in a particular position as being produced by a "stamping" or "punching" operation like a train conductor's hand puncher. When the tickets are lined up properly, it is just as easy to punch two or more tickets with the same symbol in the same position as it is to punch one ticket with that symbol.

More formally, symbols are stamped in place by a "creation machine" – some generalisation of the conductor's hand punching machine. At any intermediate stage of its manufacture, an object consists of completed stamped and uncompleted unstamped positions. The creation machine exhibits perfect economies of scale when identical stamping operations are performed on the same unstamped position of identical objects. Each identical operation on identical objects counts as only *one* operation. The appropriate image is that identical objects can be costlessly aligned so that the identical symbol may be punched in the same uncompleted position in one operation.

However, if the creation machine performs different operations on the same object or the same operation on different objects, no economies of scale are allowed and production is linear. In such cases the objects cannot be properly stacked up and more than one punching operation is required. The same operation performed on two different objects counts as two operations, just as do two different operations performed on two identical objects.

Now it seems natural to define diversity as the minimum cost of making the *n* objects different, as measured by the minimum total number of stamping or punching operations required by the creation machine. Essen-

tially, diversity is the amount of work that the creation machine must do to create the different objects.

In the case of ultrametric distances, it is easy to see that diversity equals the total branch length of the corresponding tree. In the more general case, it is possible to derive a lower bound.

The bound is derived as follows. Suppose that $Q (\subset S)$ is any set of species. Let the minimum number of operations required by the creation machine to make the collection Q be denoted $V(Q)$. Let j be a species in S but not in Q . Some reflection will reveal that the Inequality (1-3) must hold. Suppose that i is the species of Q that is closest to j . If species j is added to the collection Q , the very luckiest we might be in terms of minimising the number of operations on the creation machine is if all the positions of i that have different symbols from j happen to occur at the very end of the manufacturing sequence on the creation machine that made the collection Q , which includes i . In this fortuitous case, Equation (1-3) would hold with full equality. In the more general case, Equation (1-3) would hold as the stated inequality condition.

Now, the most optimistic number of operations of the creation machine required to make the collection S , consistent with the given distance data, must satisfy Equation (1-1). Thus, the function defined by Equation (1-1) has the interpretation of representing the minimal number of operations needed to create the diversity of the collection S .

There is yet another way of restating the evolutionary metaphor in terms of a bound on the search time for an optimal hierarchy. Here, think of each species as a library containing M books. The distance $d(i,j)$ is the number of books in library i but not in library j , or vice versa. Regularly, say once a week, every book must be searched to see if it contains some particular message, phrase, reference, formula, or whatever. It takes the same amount of time to search each book. Some books are common to two or more libraries, and the director wants to minimise the redundancy involved in searching the same book more than once. Suppose that search must be hierarchical, meaning that the catalogue must have a tree structure analogous to what is depicted in Figure 1.1. The highest level catalogue contains books held by every library. Then a bifurcation occurs which partitions the set of all libraries into two mutually exclusive subsets. The next level of (two) catalogues contains books held by all the libraries of one subset, but not by all the libraries of the other subset. Then further bifurcations occur which divide a subset of libraries into two mutually exclusive sub-subsets. Each sub-subset contains books held by all the libraries of one of the sub-subsets, but not by all the libraries of the other bifurcated sub-subset. This hierarchical catalogue process continues until all books in all libraries have been included.

In the bead model, the optimal hierarchical search procedure is identical to the genealogical tree, and it is perfectly effective in the sense that it completely avoids any redundancy of searching the same book twice in two different libraries or catalogues. When the bead model does not hold, some redundancy in hierarchical search is unavoidable. The question then becomes: what is the most economical search hierarchy in the sense of minimising total search time. Intuitively, one wants to group together libraries having a relatively large number of books in common to avoid redundant search.

When posed in full generality, it is impossible to find an optimal hierarchy without detailed information about which books are contained in every library. However, relying only on distance information does allow a theoretical lower bound on search time. This theoretical lower bound on total search time in an optimal hierarchy is what we will call the diversity of the collection of libraries.

The bound is derived as follows. Suppose that $Q (\subset S)$ is any set of libraries. Let the optimal hierarchical search time for the collection Q be denoted $V(Q)$ and the optimal search tree for Q be denoted $T(Q)$. Let j be a library in S but not in Q . Some reflection will reveal that the Inequality (1-3) must hold. Suppose that i is the library of Q that is closest to j . If library j is added to the collection Q , the very luckiest we might be in terms of minimising hierarchical search time is if all the books of i that are different from j happen to occur at the very end of the part of the search tree $T(Q)$ that involves library i . Then the optimal search tree $T(Q \cup j)$ would look just like the optimal search tree $T(Q)$ except that an extra branch of length $d(i,j)$, which contains library j as an end twig, has been appended to the branch containing library i as an end twig. In this fortuitous case, Equation (1-3) would hold with full equality. In the more general case, Equation (1-3) would hold as the stated inequality condition.

Now, following the previous logic, the most optimistic hierarchical search time consistent with the given distance data must satisfy Equation (1-1). Thus, the diversity function of this paper has the interpretation of representing the minimal amount of time needed to perform an optimal hierarchical search of all the volumes contained in a given collection of species-libraries.

Another route to forcing the Inequality (1-3) to hold so "tight" that it yields the dynamic programming Equation (1-1) is to add an extra axiom to Equation (1-3) called the "link property." This new condition can be stated as follows.

Link Property: For all S , there exists at least one species $j(S) \in S$, called the "link" species, that satisfies

$$V(S) = d(j, S|j) + V(S|j) \quad (1-6)$$

As was shown in the last section, an especially appealing theoretical structure emerges in the case of bead model distances, where, in effect, Equation (1-6) holds for *all* $j \in S$. Unfortunately, it is mathematically impossible that Equation (1-6) can be true for all $j \in S$ in the general case of non-ultrametric distances. But from the link property it will at least be true always 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 (1-3) and (1-6) imply Condition (1-1) is a fairly straightforward argument. There is also a probabilistic way of motivating the basic dynamic programming recursion of Equation (1-1) that deserves to be treated here. One of the most commonly cited reasons for maximising expected biodiversity is to maintain a kind of natural "portfolio diversification" of future options for finding new sources of food, medicine, and so forth. We will show that, under a not too bizarre model, the concept of diversity as "portfolio diversification" is really the same as the concept of diversity embodied by the diversity function previously defined.

Suppose, for concreteness, 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

$$P(i, j) \quad (1-7)$$

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 Equation (1-7) are the basic, given, reduced-form primitives of the model.

It is assumed that the given conditional probability coefficients of Equation (1-7) are symmetric; for all i and j belonging to S ,

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

Suppose the n species of S are produced by a process of "descent with modification" down an evolutionary tree, only we do not necessarily know the structure of the evolutionary tree.

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 labelled 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 has 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.

Let Q and j satisfy Equations (1-4), (1-5). 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,

$$P(j|Q) = P[j|A(Q)] \quad (1-8)$$

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,

$$P(j|A(Q)) = P[j|a(j, Q)] \quad (1-9)$$

The entire relevance for the probability that j does not contain a cure given that $A(Q)$ does not contain a cure is summarised 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,

$$P(j|i) = P[j|a(j, Q)] \cdot P(a(j, Q)|i) \quad \forall i \in Q \quad (1-10)$$

Taking the maximum of both sides of Expression (1-10) over all $i \in Q$ yields

$$\max_{i \in Q} P(j|i) = P[j|a(j, Q)] \cdot \max_{i \in Q} P(a(j, Q)|i) \quad (1-11)$$

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

$$P(a(j, Q)|k) = 1 \quad \text{for some } k \in Q \quad (1-12)$$

But then Equation (1-12) implies

$$\max_{i \in Q} P(a(j, Q)|i) = 1 \quad (1-13)$$

Combining Equation (1-13) with (1-11) with (1-9) with (1-8), and using Definition (1-7) yields

$$P(j|Q) \leq P(j, Q) \quad (1-14)$$

where

$$P(j, Q) \equiv \max_{i \in Q} P(j, i) \quad (1-15)$$

By definition

$$P(j|Q) \equiv \frac{P(Q \cup j)}{P(Q)} \quad (1-16)$$

where $P(Q)$ stands for the probability that none of the species of Q contain a cure for the disease. Combining Equation (1-14) with (1-16) yields

$$P(Q \cup j) \leq P(Q) \cdot P(j, Q) \quad \forall Q \subset S \quad \forall j \in S \setminus Q \quad (1-17)$$

Let $\Pi(S)$ stand for the *maximum* value of $P(S)$ under the set of constraints signified by Equation (1-17):

$$\Pi(S) \equiv \max P(S) \quad \text{subject to Equation (1-17)} \quad (1-18)$$

If $\Pi(S \setminus i)$ were known for all $i \in S$, then Equation (1-17) (with $j \equiv i$, $Q \equiv S \setminus i$) implies that $\Pi(S)$ defined by Equation (1-18) must satisfy the dynamic programming recursion

$$\Pi(S) = \min_{i \in S} \{\Pi(S \setminus i) \cdot P(i, S \setminus i)\} \quad (1-19)$$

It is convenient to transform (1-19) into an equivalent dynamic programming equation that is additive in distances by taking negative logarithms of all probabilities. Let

$$V(S) \equiv -\log \Pi(S) \quad (1-20)$$

$$d(i, j) \equiv -\log P(i, j) \quad (1-21)$$

Combining Equation (1-15) with (1-6) with (1-21) yields

$$d(i, S \setminus i) = -\log P(i, S \setminus i) \quad (1-22)$$

Using Equations (1-20), (1-21), (1-22), Equation (1-19) becomes transformed into the equivalent dynamic programming Equation (1-1).

From the isomorphism of Equation (1-19) with (1-1), the following important conclusion emerges. When the diversity function defined by (1-1) is being maximised, 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 minimised. Thus, provided that the distances are appropriately defined, maximising diversity is equivalent to minimising the worst-case risk of not being able to avoid some bad outcome in a portfolio choice problem. In other words, there is a not extremely bizarre model in which the two concepts of diversity are really two different sides of the same coin.

1.5 A possible application to Nature Conservancy rankings

The Nature Conservancy is a private nonprofit U.S. organisation dedicated to preserving rare or endangered species or natural communities by land acquisition programmes. In carrying out such a programme, the organisation requires operational criteria for determining site priorities.

Without going into full details, the Nature Conservancy approach to ranking the biodiversity potential of sites is very roughly the following. In Nature Conservancy methodology, the underlying "mapping units" or "elements" (species or communities) are ranked by how rare they are as measured by numbers of site occurrences: from G1 = critically important (5 or fewer occurrences) to G5 = demonstrably secure (over 100 occurrences). Then individual sites containing an element are graded by the likelihood that the element would survive on that preserved site: from A = highly likely to D = very unlikely. These two factors are then combined by prescribed guidelines to yield an overall biodiversity ranking of sites – from B1 = outstanding significance (e.g., presence of an A-ranked G1 element) to B4 = moderate significance (e.g., presence of a C-ranked G3 element). This ranking system is then used by the Nature Conservancy to prioritise the desirability of acquiring various sites.

The following question arises naturally. What is the relationship, if any, between the Nature Conservancy methodology and the diversity function approach described in this paper?

Under a number of simplifying assumptions, the Nature Conservancy biodiversity rank of a site can be interpreted as a rough approximation of the expected increase in biodiversity if the site were preserved. This may be shown as follows.

Suppose, for the sake of argument, the "elements" are species. Assume further that the distance from any species to any other species is one. (The diversity loss of each species is the same.)

Next, suppose a species i occurs on n sites. If any of these n sites is not protected, the probability of species i becoming extinct on that one site is

q . Assuming independence, the probability of species i becoming extinct in general (with loss of diversity one) is q^n .

Now suppose that one of the sites can be protected. Suppose this protection lowers the probability of extinction of species i on that site from q to p ($< q$). Then the probability that species i goes extinct in general becomes

$$pq^{n-1}$$

Defining

$$\delta \equiv q - p$$

the increase in expected diversity from protecting site i is

$$\Delta D = \delta q^{n-1}$$

In this special case, then, the change in expected diversity can be written as some function

$$f(\delta, n)$$

where f is an increasing function of δ and a decreasing function of n . In the above formula, δ is a measure of the change in site-specific survival probability or the degree of additional protection that would be given to the survival of species i on the site under consideration if that site were to be preserved.

The formula above provides a very rough justification of the Nature Conservancy ranking system. A site obtains a higher biodiversity ranking if it contains a more endangered species (n is smaller) and/or if the survival probability of the endangered species on that site is more greatly improved when the site is preserved (δ is bigger). The Nature Conservancy biodiversity ranking system corresponds very roughly to identifying sites whose preservation would cause a relatively large change in expected diversity. The Nature Conservancy methodology and the theory of diversity functions described in this paper thus dovetail quite nicely – at least under the greatly simplifying “special case” assumptions described above.

1.6 Summary and conclusions

In this paper I have tried to argue that if we are to get a handle on the conceptually very difficult problem of maximising diversity, we must be prepared to define a diversity function. Attempts to define diversity functions are in a beginning stage. I have tried to indicate one philosophy and one approach to defining diversity functions. This is not the place to develop fully the mathematical properties or the possible applications of the

diversity function being proposed.² Nor is there sufficient space to make a complete comparison between this approach and others that have been proposed in the literature.³ However, the reader should come away with at least some appreciation of the basic issues involved in defining appropriately a diversity function, and some sense of how such a definition might be constructed.

² To some extent these tasks have been attempted in my two other papers on this subject, Weitzman (1992, 1993).

³ For some examples of other approaches to diversity, see the appropriate papers cited as references.