

THE NOAH'S ARK PROBLEM

BY MARTIN L. WEITZMAN

This paper is about the economic theory of biodiversity preservation. A cost-effectiveness methodology is constructed, which results in a particular formula that can be used as a criterion to rank projects. The ranking criterion is sufficiently operational to be useful in suggesting what to look at when determining actual conservation priorities among endangered species. At the same time, the formula is firmly rooted in a mathematically rigorous optimization framework, so that its theoretical underpinnings are clear. The underlying model, called the "Noah's Ark Problem," is intended to be a kind of canonical form that hones down to its analytical essence the problem of best preserving diversity under a limited budget constraint.

KEYWORDS: Biodiversity, endangered species, diversity theory.

INTRODUCTION

THE PRESERVATION OF BIODIVERSITY is plagued by the absence of a workable cost-effectiveness framework, within which, at least in principle, basic questions can be posed and answered. Current approaches to endangered species protection seem almost completely lacking in theoretical underpinnings that might reasonably guide policy. As a result, we do not have rigorously grounded criteria for choosing among biodiversity-preserving alternatives, and it is difficult to evaluate performance.

There are several motivations for this paper. Essentially, I am trying to introduce an analytical framework that represents a useful way of thinking about the economics of diversity through the medium of an abstract mathematical model. The "Noah's Ark Problem" is a parable intended to be a kind of canonical form of the simplest possible way of representing how best to preserve biodiversity under a limited budget constraint. Throughout the paper, emphasis is laid upon simplicity of structure, in order to get at the analytical essence of the problem.

While several themes are developed in this paper, the main motivating question is how to determine basic priorities for maintaining or increasing diversity. The central issue is to develop a cost-effectiveness formula or criterion that can be used to rank priorities among biodiversity-preserving projects. The formula should be operational enough to be useful in suggesting what to look at when actually determining conservation priorities, e.g., among endangered species. At the same time, the methodology should be firmly rooted in a mathematically rigorous optimization model, so that the theoretical foundation for the cost-effectiveness ranking criterion is made very clear.

In talking about biodiversity preservation, there is always a question about what is the appropriate level of discourse. In principle, the basic unit could be at

the level of the molecule, cell, organ, individual, species, habitat, ecosystem, or so forth. For the purposes of this paper, I take the underlying unit of analysis to be the “species,” although I think that the same basic issues and themes of the paper will arise at any level. In this spirit, the Noah’s Ark Problem could be thought of as a generic formulation of “the” core problem of optimal biodiversity preservation under a budget constraint.

Suppose, then, that the underlying conservation unit i essentially stands for “species i .” In the context of this paper, it is useful to conceptualize a “conservation project” as follows. “Project i ” is some preservation action that increases the probability of survival of species i by ΔP_i at a cost of C_i . Let U_i represent the direct utility of how much we like or value the existence of species i . (After all, most of us like Pandas more than mosquitoes.) Without trying to be precise about it at this stage, let the “distinctiveness” of i , meaning roughly its difference or distance from its closest resembling unit, be D_i . (Imagine somehow that D_i has been made commensurate with U_i .) Then the following formula might appear on heuristic grounds to be conveying roughly the right priorities for ranking alternatives:

$$(1) \quad R_i = (D_i + U_i) \left(\frac{\Delta P_i}{C_i} \right).$$

As a ranking criterion, R_i is a measure of the “expected marginal distinctiveness plus utility per dollar.” Interpreted loosely, when making preservation decisions the conservation authorities are asked to look at the four factors: D_i , U_i , ΔP_i , C_i , which seems reasonable, combined in such a way as described by formula (1), which also seems intuitively plausible. The proposed formula has some heuristic appeal as a rule of thumb, and it seems to make sense generally, but can it be grounded in any theoretical construct? Does the ranking criterion (1) have a formal justification?

The primary task of the paper is to present a model and a scenario where the use of R_i as a ranking criterion can be given a rigorous underpinning. The main question I am asking is this: If the “right assumptions” are made, can a reasonable optimizing model be produced that provides a rigorous justification for using R_i to select conservation priorities? The short answer is “yes,” and the underlying model that provides the “right assumptions” I call the “Noah’s Ark Problem.”

Trying to answer the above, rather narrowly posed, question about an appropriate ranking criterion forces a researcher to confront a number of basic theoretical and conceptual issues that go to the core of modeling diversity preservation. My hope is that the basic conceptual framework and the set of ideas introduced here for thinking about the economics of diversity will prove useful even beyond the specific applications of this paper.

The model itself will be at a high level of abstraction. My aim is to build the simplest possible analytical structure that captures the mathematical essence of the underlying problem. In so doing, I entertain the standard economic theorist’s

hope that basic insights into policy which emerge from the bare-bones canonical-form model will remain a useful point of departure even in a more complicated world.

THE "SPECIES / LIBRARY" MODEL OF DIVERSITY

A key point of departure for this paper involves conceptualizing the underlying conservation unit—the "species"—as if it were a "library." Concentrating on the question of how best to preserve libraries allows for a crisp formulation of the generic problem of optimally conserving diversity under a budget constraint. In this section, then, the basic unit is the library. Each individual library stands, at an appropriate level of abstraction, for a particular species.

Let the index $i = 1, 2, \dots, n$ denote a particular library. The set of all n libraries is denoted S .

The value of a library consists of two components: the building itself and the collection of books that it houses. Library i is housed in a building that has some inherent value as a structure—let this utility be denoted U_i . The other, presumably primary, value of library i is its book collection—suppose this consists of M_i different books. (In the species interpretation, U_i represents the direct utility of how much we like or value the existence of species i , while the M_i "books" are very roughly analogous to the gene pool of the species itself.)

The book collections in various libraries may overlap to some degree. The definition of the *diversity of S* , denoted $V(S)$, is the number of *different* books in the overall library collection S . In other words, the *diversity function $V(S)$* measures the size of the set that consists of the union of all the different books in all the libraries of S .

A simple two-library example may help to illustrate the basic idea. Let S consist of the two libraries 1 and 2. Let the number of books held jointly in common between libraries 1 and 2 be denoted J . Let E_1 stand for the number of books distinctive or unique to library 1, while E_2 is the number of books distinctive to library 2. Then $M_1 = E_1 + J$, while $M_2 = E_2 + J$.

In the case of two libraries, it is always possible to give a tree representation of the book structure. The tree, depicted in Figure 1, has a main branch of length J that divides into two sub-branches of lengths E_1 and E_2 . (One could tell an evolutionary story "as if" the two libraries evolved by "descent with modification" from a common ancestor.) The diversity function in this case is $V(S) = J + E_1 + E_2$. Note that diversity here has the interpretation of being the total (vertical) branch length of the corresponding tree.

We defined diversity for the particular assemblage S of libraries. But essentially the same idea holds for *any* assemblage of libraries. For any subset of libraries Q ($\subseteq S$), define the *diversity of Q* , denoted $V(Q)$, to be the number of *different* books in all the libraries of Q . In other words, the *diversity function $V(Q)$* is the size of the set consisting of the union of all books in the libraries of Q , where Q may represent essentially any assemblage of libraries.

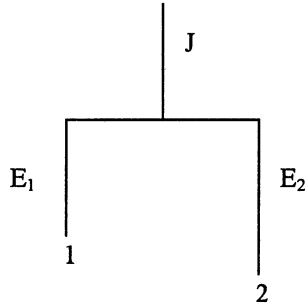


FIGURE 1.

A critical aspect of the underlying problem is uncertainty. To continue the metaphor, a library can burn down, with possible loss of the building and the book collection that it houses. Various preventive measures can be undertaken that lower the probability of a fire—at a cost. The essence of the conservation problem is how best to allocate scarce fire prevention resources among the various libraries.

Uncertainty is introduced as follows. Let P_i be the probability of survival of library i , while $1 - P_i$ is the probability of extinction of i —e.g., the probability that library i burns down. For analytical simplicity, I assume throughout the paper that all probabilities are *independent*, since this is the easiest case to analyze.

As previously noted, the direct utility of library building i is U_i , which represents how much the library building itself is liked irrespective of its book contents. Without loss of generality, suppose that the coefficient U_i is normalized relative to the value of each different book, which is set equal to one. Then the expected value of all the library buildings is

$$(2) \quad U(P) = \sum P_i U_i.$$

Turning now to the book collections, the definition of expected diversity is a natural extension of the original deterministic concept to the case of uncertainty. The appropriate stochastic concept here is the probability-weighted diversity of all subsets. The *expected diversity function*, denoted $W(P)$, is the *average number of different books*. More formally,

$$(3) \quad W(P) \equiv \sum_{Q \subseteq S} \left(\prod_{j \in Q} P_j \right) \left(\prod_{k \in S \setminus Q} (1 - P_k) \right) V(Q).$$

In the two-library example,

$$(4) \quad W(P) = P_1 P_2 \cdot (M_1 + M_2 - J) + P_1 (1 - P_2) \cdot M_1 + P_2 (1 - P_1) \cdot M_2 + (1 - P_1)(1 - P_2) \cdot 0$$

which can be rewritten more simply as

$$(5) \quad W(P_1, P_2) = M_1P_1 + M_2P_2 - JP_1P_2.$$

Note that $U(P)$ represents the expected utility of the library *buildings*, while $W(P)$ represents the expected value of the overall collection of *books*. Of the two functions, $W(P)$ is by far the more novel or unorthodox, and the really innovative part of this paper deals with it.

The expected diversity function $W(P)$ expresses the average number of *different* books in the collection. Now we might like many different books per se, just as we might like many different colors simply because of the more colorful world their sheer diversity creates. This would be a kind of *aesthetic value* of diversity. Or, we might want to have different books for the utilitarian reason that they are a potential source of new future ideas about medicines, foods, or whatever. This might be called the *information content* of a book collection. These two concepts are actually closely related to each other. As the next section shows, at a sufficiently high level of abstraction the *aesthetic value* of diversity is essentially the same concept as the *information content* of diversity.

EXPECTED DIVERSITY AS INFORMATION CONTENT

This section develops a strict isomorphism between “diversity as aesthetic value” and “diversity as information content.” It will be proved that when they are appropriately modeled, the two concepts are formally identical.

Suppose we take the average number of books $W(P)$ as an appropriate measure of the aesthetic value of diversity. What should be the appropriate measure of the information content of diversity?

A “book” in the present model is a metaphor for a container of information. Suppose that a researcher is looking for a particular piece of useful information. There are three possibilities: the information might not exist in any of the books; or it might be located in just one book; or it might be redundantly contained in more than one book. Suppose that the probability that any one book contains this particular piece of useful information is ε . A vivid image is that ε stands for the probability that in any book drawn at random is contained the alchemist’s secret of how to turn lead into gold. In a more realistic scenario, the “useful information” being sought might be about a particular cure for a certain disease or a specific new food source.

Consider any assemblage of libraries Q . Then the probability that Q contains useful information is the probability that *at least one book* of Q contains the useful information. Since there are $V(Q)$ different books in Q , this equals one minus the probability that *all* $V(Q)$ books fail to contain the useful information. Thus, the probability that Q contains useful information is

$$(6) \quad H(Q; \varepsilon) \equiv 1 - (1 - \varepsilon)^{V(Q)}.$$

Note that taking the derivative of (6) with respect to ε and evaluating at $\varepsilon = 0$ yields the relationship

$$(7) \quad \left. \frac{\partial H(Q; \varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} = V(Q).$$

The expression (6) for $H(Q; \varepsilon)$ represents the probability that the *deterministic* library assemblage Q contains useful information. But under uncertainty, the library assemblages themselves exist only in a probabilistic sense. Library i exists with probability P_i . The library assemblage $Q (\subseteq S)$ exists with probability

$$(8) \quad \left[\prod_{j \in Q} P_j \right] \cdot \left[\prod_{k \in S \setminus Q} (1 - P_k) \right].$$

It is natural to define formally the *information content* I (as a function of P and ε) to be the probability that the overall stochastic library collection contains useful information on a particular topic. Metaphorically, information content is just the probability that in at least one library that has not burned down is contained at least one book in which is written the alchemist’s secret of how to turn lead into gold.

Combining (6) with (8), the *information content* of the stochastic library collection is

$$(9) \quad I(P; \varepsilon) = \sum_{Q \subseteq S} \left(\prod_{j \in Q} P_j \right) \left(\prod_{k \in S \setminus Q} (1 - P_k) \right) [H(Q; \varepsilon)].$$

What is the relationship between *diversity* and *information content*? Essentially, the basic result here is that diversity is the first derivative of information content with respect to uncertainty. Taking the derivative of (9) with respect to ε , using (7) to evaluate it at $\varepsilon = 0$, and comparing with definition (3) yields the basic result:

$$(10) \quad \left. \frac{\partial I(P; \varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} = W(P).$$

Since it is readily verified that $I(P; 0) = 0$, equation (10) means that $W(P) \cdot \varepsilon$ is the appropriate first order approximation of $I(P; \varepsilon)$ in the formal sense that

$$(11) \quad I(P; \varepsilon) = W(P) \cdot \varepsilon + O(\varepsilon^2)$$

where $O(\varepsilon^2)$ stands for squared or higher order terms in ε .

The conclusion is that, for small ε , maximizing the expected diversity function $W(P)$ is “essentially” the same as maximizing information content $I(P; \varepsilon)$.

We state this result somewhat informally as the following:

THEOREM 1: *The expected diversity of a set of library collections is “essentially” the same concept as the information content of the same set of library collections.*

What do we want to do with expected diversity (equals information content)? Presumably we want to pick conservation policies that maximize the sum of the expected diversity function of the books $W(P)$ plus the expected utility of the library buildings $U(P)$, taking account of the relevant budget-like constraints on P . The simplest canonical form of a preservation problem involving optimizing under budget constraints, which at the same time conveys some useful content, is what I call here the “Noah’s Ark Problem.”

THE NOAH’S ARK PROBLEM

The “Noah’s Ark Problem” is intended to be an allegory or parable that renders a vivid image of the core problem of maximizing diversity under a budget constraint. What is treated here is actually not the most general form of the underlying mathematical problem. Some slight generalizations are possible, but they would come at the expense of diluting a crisp version of the basic paradigm.

Noah knows that a flood is coming. There are n existing species/libraries, indexed $i = 1, 2, \dots, n$. Using the same notation as before, the set of all n species/libraries is denoted S . An Ark is available to help save some species/libraries. In a world of unlimited resources, the entire set S might be saved. Unfortunately, Noah’s Ark has a limited capacity of B . In the Bible, B is given as $300 \times 50 \times 30 = 450,000$ cubits³. More generally, B stands for the total size of the *budget* available for biodiversity preservation.

In either case, Noah, or society, must face the central problem of *choice*. A basic choice question must be answered. Which species/libraries are to be afforded more protection—and which less—when there are not enough resources around to fully protect everything? I present here the simplest way that I know to model the analytical essence of this choice problem.

If species/library i is *not* put on board the Ark, but left unprotected, its probability of surviving the flood unaided is \underline{P}_i . If species/library i is boarded on the Ark, and thereby afforded some protection, its survival probability is enhanced to \overline{P}_i . Essentially, boarding on the Ark is a metaphor for investing in a conservation project, like habitat protection, that improves survivability of a particular species/library.

A particularly grim version of the Noah’s Ark Problem would make the choice a matter of life or death, meaning that $\underline{P}_i = 0$ and $\overline{P}_i = 1$. This specification is perhaps closest to the old testament version, so I am taking literary license here by extending the metaphor to less stark alternatives. The only mathematical restriction here is then

$$(12) \quad 0 \leq \underline{P}_i \leq \overline{P}_i \leq 1.$$

There is also a vital “cost” side to the problem. It is natural to suppose that if species/library i is boarded, it takes up some space or room on the Ark—let this space coefficient be denoted C_i . Since overall space is limited, the amount

of room that a species occupies becomes a critical factor in the choice problem. For the biblical Ark parable, C_i is measured in units of cubits³. In the real world, C_i would represent the *cost* of the project that extends an enhanced measure of protection to species/library i .

We can now state the central problem. Noah wishes to select survival probabilities that maximize expected diversity plus expected direct utility. The Noah's Ark mathematical programming problem is to select values of $\{P_i\}$ that

$$(13) \quad \text{maximize}_{\{P_i\}} [W(\{P_i\}) + U(\{P_i\})]$$

subject to the n individual probability constraints

$$(14) \quad \underline{P}_i \leq P_i \leq \overline{P}_i \quad \forall i,$$

and subject to the overall budget constraint

$$(15) \quad \sum C_i \left(\frac{P_i - \overline{P}_i}{\overline{P}_i - \underline{P}_i} \right) = B.$$

Note that, as formulated, the above programming problem is continuous in the probability variable P_i , and thus allows for strictly "interior" values of P_i , which fulfill the condition

$$(16) \quad \underline{P}_i < P_i < \overline{P}_i.$$

An "interior" value of P_i , one that satisfies (16), corresponds to partial protection, or "fractional boarding" of species/library i on the Ark. "Fractional boarding" might be given a physical interpretation of boarding only some fraction of a reproductively viable population size. Suppose that, when not a single individual of species/library i is boarded, the survival probability would be $P_i = \underline{P}_i$. At the opposite extreme, if population size C_i is boarded, then suppose the survival probability is enhanced to $P_i = \overline{P}_i$. Finally, assume that for "in-between" population size boardings a linear interpolation describes the survival probabilities. Under such assumptions, there is a rigorous underlying basis for saying that a survival probability of P_i for species/library i comes at a budget cost of

$$(17) \quad C_i \left(\frac{P_i - \underline{P}_i}{\overline{P}_i - \underline{P}_i} \right).$$

I now seek to characterize the nature of the solution to the Noah's Ark Problem. The most distinctive aspect of the mathematical programming problem (13)–(15) is the presence of the unusual "expected diversity function" $W(P)$ in the objective. It might be thought that the function $W(P)$ is so unorthodox that it is difficult to say anything general that is also interesting about the solution to the problem. Fortunately, it turns out that a quite striking characterization is possible.

The solution of the Noah's Ark Problem is always "extreme" in the following sense. Noah, or the conservation authorities that he symbolizes, should be concentrating all their resources on maximal protection of some selected species/libraries, even at the expense of exposing all remaining species/libraries to minimal protection.

In an optimal policy, the entire budget is spent on a favored subset of species/libraries that is afforded maximal protection. The less favored complementary subset is sacrificed to a level of minimal protection in order to free up to the extreme all possible scarce budget dollars to go into protecting the favored few.

While a real-world interpretation of this result must be properly qualified, because it is only as strong as the underlying assumptions, the following thought leaps to mind. Subject to the restrictions of the model, there is some implication that a conserving agency may want to think more in terms of concentrating limited resources, rather than spreading them out thinly.

In order to state the above ideas rigorously, we start with the following definition.

DEFINITION (Extreme Policy Solution): "Almost all" ($n - 1$ out of n) species/libraries are either fully boarded ($P_i = \bar{P}_i$) or not boarded at all ($P_i = \underline{P}_i$). At most one species/library j (the "roundoff species/library") is "fractionally boarded" with interior probability P_j , where $\underline{P}_j < P_j < \bar{P}_j$.

The basic result here is the following:

THEOREM 2: *The solution of Noah's Ark Problem is an extreme policy.*

PROOF: I present here a concise version, which the reader should be able to follow even though it is compact. Not every single aspect is spelled out, because to do so requires a lot of algebra and notational detail, which the interested reader should be able to fill in. All the main steps in the underlying logic are provided.

The proof begins with the observation that conditioning on the existence of species/libraries 1 and 2 allows the expected diversity function $W(P)$ to be rewritten in the form:

$$(18) \quad W(P) = \sum_{Q \subseteq S \setminus \{1,2\}} [K(Q)] \cdot [P_1 P_2 V(Q \cup \{1,2\}) + P_1(1 - P_2)V(Q \cup \{1\}) \\ + P_2(1 - P_1)V(Q \cup \{2\}) + (1 - P_1)(1 - P_2)V(Q)]$$

where $K(Q)$ is defined as the polynomial expression:

$$(19) \quad K(Q) \equiv \left(\prod_{j \in Q} P_j \right) \cdot \left(\prod_{k \in S \setminus Q \setminus \{1,2\}} (1 - P_k) \right) \geq 0$$

and set notation such as $Q \cup \{1\}$ or $Q \setminus \{1\}$ are shorthand for, respectively, $Q \cup \{1\}$ or $Q \setminus \{1\}$.

Mechanically taking the second mixed partial derivative of (18) yields, after some algebraic manipulation, the formula

$$(20) \quad \frac{\partial^2 W}{\partial P_1 \partial P_2} = \sum_{Q \subseteq S \setminus \{1,2\}} -[K(Q)] \cdot [J(Q)]$$

where $J(Q)$ is just shorthand notation for the following expression:

$$(21) \quad J(Q) \equiv [V(Q \cup 1) + V(Q \cup 2) - V(Q \cup 1 \cup 2) - V(Q)].$$

Now the $J(Q)$ formula expressed by (21) actually stands for something. Looking carefully at the appropriate Venn diagram should convince the reader that $J(Q)$ stands for the number of books held *jointly* by libraries 1 and 2 but *not* contained in any *other* libraries of Q . Naturally, $J(Q)$ is nonnegative.

Thus, the expression (20) for the second mixed partial of W with respect to P_1 and P_2 is nonpositive and independent of P_1 or P_2 . It follows that the expected diversity function $W(P)$ is convex in *any two* of its variables, holding all other arguments constant.

The objective function (13) is therefore convex in any two of its variables, holding all other variables constant, while the relevant constraints (14) and (15) are linear. Because the optimal value of a convex function maximized over a convex set is always attained at an extreme point of the convex set, it follows that, out of *any pair* of probability variables in an optimal policy, no more than one would be strictly “interior” in the sense of satisfying condition (16). But this means that *at most one* species/library j , out of the entire set S , is “fractionally boarded” with interior probability P_j , where $\underline{P}_j < P_j < \bar{P}_j$. Q.E.D.

The truly “extreme” nature of this solution can perhaps best be appreciated by setting $\underline{P}_i = 0, \bar{P}_i = 1$ for all i . Then a diversity-maximizing policy makes each species (except for at most one) either totally extinct or perfectly safe.

It is important to understand the intuitive logic that explains why an extreme policy is optimal in the Noah’s Ark Problem. Consider the two-library example depicted in Figure 1. As before, let E_1 be the number of books distinctive to library 1 and E_2 the number of books distinctive to library 2, while the number of books held jointly in common between libraries 1 and 2 is denoted J . Then the relevant expression for expected diversity in this case is given by (5).

If library 1 *alone* becomes extinct, then E_1 books are lost. If library 2 *alone* becomes extinct, then E_2 books are lost. However—and this is the crucial point—if libraries 1 and 2 *both* become extinct, then an *additional* J books are lost. With costs linear, Noah would rather shift probabilities toward saving as fully as possible one of the libraries at the expense of the other in order to “pin down the line” of the J books in common. The strength of this effect of increased returns to saving either one of the libraries is measured by J , which from (5) is minus the second mixed partial derivative of expected diversity, corresponding to equation (20).

The idea of “*pinning down the line of books in common*” seems to me to be an important conservation principle. In the case of costs that are linear in probabilities, it results in policies of extreme concentration of conservation resources. As the incremental cost of protection increases nonlinearly in survival probability, this extremeness of an optimal policy is diluted, or even reversed for sufficiently strong curvature. Nevertheless, the basic principle remains, and may even give useful insight into the form of an optimal conservation policy under some circumstances. When costs are approximately linear in probability changes, as might be supposed to be the case for sufficiently small $\Delta P_i \equiv \bar{P}_i - \underline{P}_i$, then Theorem 2 articulates a well-defined sense in which a conserving agency should be concentrating scarce resources on some libraries for the sake of “pinning down the line of books in common” that might otherwise be lost from multiple extinctions.

Theorem 2 of this section gives a strong characterization of the form of a solution to the Noah's Ark Problem. The solution form is an extreme policy where the preservation changes of some species/libraries are maximally enhanced by their being boarded completely on the Ark, while others are left completely behind. But this result does not say *which* species/libraries are favored by being selected for boarding, nor explain fully *why* they are so favored.

As a means towards the end of actually solving the Noah's Ark Problem, there is a need to understand better the properties of the expected diversity function and also to incorporate into the analysis some measure of uniqueness or distinctiveness of a species/library. It turns out that these two issues are intimately related. There exists a striking connection between what might be called the “marginal diversity” of a species/library and what might be called its “distinctiveness,” meaning loosely a distance-based measure of difference from other species/libraries. To bring out this important relation, some definitions need to be made and some more structure must be put on the problem. The additional structure that will eventually be imposed on species/libraries is that they are “as if” created by a specific process of descent with modification, which will be called the “Evolutionary Library Model.”

“DISTINCTIVENESS” IN THE EVOLUTIONARY LIBRARY MODEL

We now wish to define the uniqueness or “distinctiveness” of a library. Essentially, we identify here “distinctiveness” with the fundamental mathematical concept of distance. Consequently, there is a need to start with the relevant measure of distance in this setting. The definition of pairwise species/library distance appropriate to the present context is as follows:

DEFINITION: *Distance* $D(i, j)$ is the number of different books contained in library i that are not contained in library j .

For the important special case where libraries i and j contain the same number of different books, distances are symmetric because $D(i, j) = D(j, i)$

whenever $M_i = M_j$. It is important to note, however, that distances as defined above are not symmetric in the general case where M_i is not equal to M_j .

Now let Q be any assemblage of libraries. In the present context it is appropriate to employ the standard mathematical definition of the distance from a point to a set:

$$(22) \quad D(i, Q) \equiv \min_{j \in Q} D(i, j).$$

In the case of Q being the null set, distances are normalized so that $D(i, \emptyset) = M_i$.

The interpretation of $D(i, Q)$ is that it represents the distance of library i from its "nearest neighbor" or "closest relative" in Q .

It seems natural to define the *distinctiveness* of i , denoted D_i , to be the *expected distance* of i from its nearest neighbor or closest relative in S :

$$(23) \quad D_i \equiv \sum_{Q \subseteq S \setminus i} \left(\prod_{j \in Q} P_j \right) \left(\prod_{k \in S \setminus Q \setminus i} (1 - P_k) \right) D(i, Q).$$

What is the possible connection between distinctiveness and diversity? To answer this question sharply, one needs to put more structure on the problem. I now try to impose a "natural" structure on the book collections of the various species/libraries, which corresponds, at a high level of abstraction, to the standard paradigm of evolutionary relationships among biological species. The book collections of the various libraries will be modeled "as if" they were acquired by a process of "descent with modification."

Previously, no restrictions were placed upon the book collections of the various libraries. Now I want to suppose that the book collections are "as if" they were acquired by an evolutionary branching process with a corresponding evolutionary tree structure. The particular branching process described here is called the *evolutionary library model*, and it is patterned on the classic paradigm of biological species evolution.

The "evolutionary library model" is a branching process that explains the existence of the current library assemblage S as a result of three types of evolutionary/historical events.

(i) Each existing library acquires new books at any time by *independently sampling*, at its own rate, out of an infinitely large pool of *different* books. The independent acquisition of different new books by each library corresponds to the evolution of genetic traits when species are reproductively isolated with no gene pool mixing by lateral transfer.

(ii) New libraries can be created by a "speciation event." A new branch library can be founded by adopting a complete copy of the current collection of an existing library. Henceforth this new library will become reproductively isolated and acquire its books independently, as described by (i) above.

(iii) Libraries can go extinct. When a library is extinguished, its entire collection of books is lost. Libraries that have already gone extinct in the past

correspond to lost stem taxa, and do not show up in the set S of currently existing libraries.

Figure 2 illustrates the evolutionary library model for the case of three species/libraries. The first two libraries correspond to the depiction of Figure 1, shown in a previous section of the paper. As was pointed out there, any two libraries can be given an evolutionary tree representation. But for three or more libraries, the evolutionary library model must be assumed in order to have a tree representation. Hence, three libraries is the simplest case to analyze where the evolutionary library model is actually imposing additional structure on the library collections.

The evolutionary library story told by Figure 2 is something like the following. At the beginning there was one prototype library, which acquired a collection of G different books. Then occurred a "speciation event." Two "reproductively isolated" libraries were created, each starting off with identical copies of the G different books. Both libraries then began independently to acquire different books. One of the two libraries eventually became, after acquiring independently E_3 different books, the currently existing library 3. The other library was the ancestor of libraries (1, 2). This ancestor library of (1, 2) acquired independently J different books before another "speciation event" occurred—resulting in the fission-like founding of two identical libraries. One of these libraries, after further independent acquisition of E_1 different books, became eventually library 1. The second library, after further acquiring independently E_2 different books, evolved into library 2. In this way did the current collection of libraries {1, 2, 3} come into being.

I present here the "evolutionary library model" in a particularly sharp form to emphasize its essential characteristics. As such, the concept is intended to represent an "ultimate abstraction" of how entities are created by an evolutionary branching process. While the model could tolerate some generalizations,

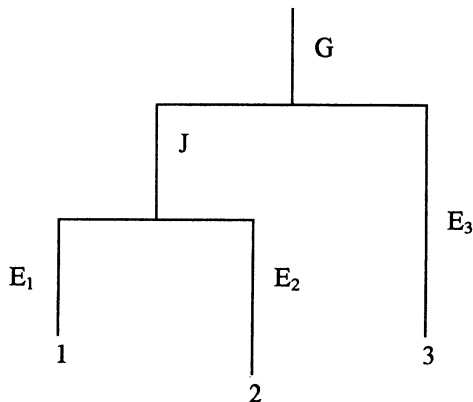


FIGURE 2.

such as types of nonindependent sampling, I do not feel it is worth the loss of sharp focus to pursue here the most general possible formulation.

In thinking about the evolutionary library model, it is almost as if the fundamental units are *defined* as entities that satisfy the three basic axioms. Since the model is at such a high level of generalization, in a way it does not really matter whether “evolutionary libraries” stand for biological species, or languages, or something else. An “evolutionary library” here is an abstraction standing for an entity that evolves as a unit independently of other such units, and which came into being originally by splitting off from another such entity. I think there is some merit in first defining carefully the mathematical essence of an evolutionary branching process—i.e., the “evolutionary library model”—and then, in the biological context, *defining* species to be units that satisfy, not perfectly but tolerably well in practice, such a model.

The evolutionary library model naturally generates a corresponding evolutionary tree. And when a tree structure is present, it seems to induce naturally in the human mind a way of visualizing and comprehending intuitively relationships among objects that are quite subtle or complicated to describe without the tree. “Tree thinking” represents a prime example of how one picture may be worth a thousand words.

Take, for example, diversity, which is defined as the total number of different books. In the evolutionary library model, diversity has a natural interpretation as the total (vertical) branch length of the corresponding tree. For the three-library example of Figure 2, it is readily confirmed that the diversity of the three libraries is the total branch length $G + J + E_1 + E_2 + E_3$.

How is “distinctiveness” represented in the context of an evolutionary branching model? Absent uncertainty, the distinctiveness of a library is just its distance from its nearest neighbor or closest relative. In the tree corresponding to an evolutionary branching model, the distinctiveness of a library is represented by its (vertical) branch length off the rest of the evolutionary tree. Thus, in Figure 2, one readily confirms that the distinctiveness of library i is represented by its (vertical) branch length E_i for all $i = 1, 2, 3$.

Note here an important geometric relationship. For an evolutionary library in the deterministic case, the loss of diversity from extinction is just the length of the (vertical) library branch from the rest of the evolutionary tree. The precise mathematical statement is

$$(24) \quad V(Q \cup i) - V(Q) = D(i, Q).$$

A vivid image is that the extinction of a library corresponds to snapping off its branch from the rest of the evolutionary tree. In Figure 2, the loss of diversity if library 1 goes extinct is E_1 . The loss of diversity if library 2 goes extinct is E_2 . (The loss of diversity if libraries 1 *and* 2 *both* go extinct is $E_1 + E_2 + J$, which is the total (vertical) branch loss of sequentially snapping off 1 first and then 2 next, or vice versa.)

In the deterministic case, the loss of diversity from extinction of a library exactly equals the distinctiveness of the library. Essentially, change of diversity is

distinctiveness. This important relation generalizes from a deterministic to a probabilistic situation. The basic idea can be stated formally as the following result:

THEOREM 3: *In the case of evolutionary libraries,*

$$(25) \quad \frac{\partial W}{\partial P_i} = D_i \quad \text{for all } P_i \in [0, 1].$$

PROOF: Rewrite the diversity function (3) as

$$(26) \quad W(P) \equiv P_i \sum_{Q \subseteq S \setminus i} \left(\prod_{j \in Q} P_j \right) \left(\prod_{k \in S \setminus Q \setminus i} (1 - P_k) \right) V(Q \cup i) \\ + (1 - P_i) \sum_{Q \subseteq S \setminus i} \left(\prod_{j \in Q} P_j \right) \left(\prod_{k \in S \setminus Q \setminus i} (1 - P_k) \right) V(Q).$$

Taking the derivative of (26) with respect to P_i and collecting terms yields

$$(27) \quad \frac{\partial W}{\partial P_i} = \sum_{Q \subseteq S \setminus i} \left(\prod_{j \in Q} P_j \right) \left(\prod_{k \in S \setminus Q \setminus i} (1 - P_k) \right) [V(Q \cup i) - V(Q)].$$

Combining (27) with (24) and (23) yields the result (25).

Q.E.D.

Theorem 3 is a statement that marginal diversity and distinctiveness are the same concept in the context of an evolutionary library model. The theorem represents an appropriate generalization to the uncertain case of the essential idea from the deterministic case that loss of diversity can be visualized as a branch-snapping event.

There is, however, a new wrinkle added by the presence of uncertainty. Distinctiveness now is associated with *expected* distance from a nearest neighbor or closest relative, loosely speaking. Under uncertainty, seemingly symmetric libraries can have different degrees of distinctiveness, with consequences for conservation policy. It is important to understand how this comes about.

To see this effect most starkly, return to the simple two-library example of Figure 1. Make the tree picture symmetric by supposing that both libraries have the same number of books so that $M_1 = M_2 = M$, and also $E_1 = E_2 = E$. Taking the derivative of expression (5), and using relation (25), yields that $D_1 = M - JP_2$, while $D_2 = M - JP_1$. Thus, in a seemingly symmetric situation, the *more* endangered species is *less* distinctive.

To expose sharply the underlying logic behind this seemingly counterintuitive result, suppose an extreme example where $P_1 = .99$, while $P_2 = .01$. Then library 1 is practically safe, while library 2 is practically extinct. For such a situation, library 2 is much less distinctive than library 1 in the following sense. The presence of library 2 is practically contributing only E different books, since library 1 is almost sure to survive by itself. On the other hand, the presence of

library 1 is practically contributing $E + J$ different books, because library 2 will almost surely become extinct. In this sense library 1 is much more distinctive than library 2. The principle of “*pinning down the line*” of jointly held books manifests itself in this example by indicating the relatively safe library as more distinctive than the relatively endangered library. Pushing even further the extreme symmetry assumptions, if the underlying costs of changing probabilities are the same, then it is optimal to make the safe library even safer at the expense of making the endangered library even more endangered—because the safe library is more distinctive and has greater marginal diversity.

At this point all of the necessary analytical apparatus has been developed, and the paper is ready to begin its main theme of developing a ranking criterion that solves the Noah’s Ark Problem.

THE NOAH’S ARK MYOPIC RANKING CRITERION

Suppose Noah wishes to actually solve problem (13)–(15). He wants to maximize expected diversity plus direct utility subject to the relevant constraints. He does not want, however, to mess around with a complicated algorithm. Noah is a practical outdoors man. He needs robustness and rugged performance “in the field.” As he stands at the door of the ark, Noah desires to use a simple priority ranking list from which he can check off one species at a time for boarding. Noah wishes to have a robust rule in the form of a basic ordinal ranking system so that he can board first species #1, then species #2, then species #3, and so forth, until he runs out of space on the ark, whereupon he battens down the hatches and casts off.

Can we help Noah? Is the concept of an ordinal ranking system sensible? Can there exist such a simple myopic boarding rule, which correctly prioritizes each species *independent* of the budget size? And if so, what is the actual formula that determines Noah’s ranking list for achieving an optimal ark-full of species?

The answer to these questions is essentially positive, and along the following lines.

Provided that

$$(28) \quad \Delta P_i \equiv \bar{P}_i - P_i$$

is “relatively small”¹ (for all i) in the usual sense of the prototypical small project justifying cost-benefit investment methodology locally, then a priority ranking based on the criterion

$$(29) \quad R_i = [D_i + U_i] \left(\frac{\Delta P_i}{C_i} \right)$$

¹ Note that the presumed smallness of ΔP goes somewhat against the spirit of the biblical version of Noah’s Ark, for which a fair interpretation might be $\Delta P = 1$. If so, I plead literary license to justify using the extended metaphor here, because it is so pretty.

is justified in the sense of giving an arbitrarily close first order approximation to an optimal policy.

To intuit why this might be so, ask the following question. If we have enough money to adjust probabilities a little bit in a particular direction, which direction would you choose? This is asking about the “gradient” of the objective function. Theorem 4 says that the gradient indicates a derivative of $D_i + U_i$ in the direction of P_i , which implies a policy of the specified extreme form that pushes probabilities to their maximum or minimum value.

More formally, we have the following theorem.

THEOREM 4: *Maintain the evolutionary-library hypothesis. Suppose one selects a solution of the Noah’s Ark Problem to be of the following form: There exists a cutoff value R^* such that*

$$R_i > R^* \Rightarrow P_i = \overline{P}_i \quad (\text{species } i \text{ is boarded}),$$

$$R_i < R^* \Rightarrow P_i = \underline{P}_i \quad (\text{species } i \text{ is not boarded}).$$

Then the error introduced by this proposed solution is of second or higher order in $\{\Delta P_i\}$.

PROOF: The proof presented here is concise. Not every single aspect is spelled out, since to do so would require a lot more algebra and notation. While some of the messier details about the application of Taylor-series approximations are left to the reader, all of the main steps in the underlying logic are provided here.

For all $\{P_i\}$ satisfying (14),

$$(31) \quad W(\{P_i\}) \approx W(\{\underline{P}_i\}) + \sum_{i=1}^n \left[\frac{\partial W}{\partial P_i} \right]_{P_i=\underline{P}_i} (P_i - \underline{P}_i),$$

where the symbol \approx stands for a first order approximation, which is arbitrarily accurate for sufficiently small $\{\Delta P_i\}$ in the traditional sense of a Taylor-series expansion that omits only terms of order $(\Delta P_i)^2$ or higher.

But from Theorem 3,

$$(32) \quad \left[\frac{\partial W}{\partial P_i} \right]_{P_i=\underline{P}_i} = D_i.$$

Using (32), rewrite (31) as

$$(33) \quad W(\{P_i\}) \approx W(\{\underline{P}_i\}) + \sum_{i=1}^n D_i (P_i - \underline{P}_i).$$

Now substitute the linearized expression (33) into the objective function (13), for which it is an arbitrarily close approximation. The reduced form of the

linearized version of (13)–(15) is now a programming problem that selects values of $\{P_i\}$ that

$$(34) \quad \text{maximize}_{\{P_i\}} \sum \alpha_i P_i + \text{constant}$$

subject to the n individual probability constraints

$$(35) \quad \underline{P}_i \leq P_i \leq \overline{P}_i, \quad \forall i,$$

and subject to the overall budget constraint

$$(36) \quad \sum \beta_i P_i = \gamma.$$

In the linearized problem (34)–(36), the following definitions are employed:

$$(37) \quad \alpha_i \equiv D_i + U_i$$

and

$$(38) \quad \beta_i \equiv \frac{C_i}{\Delta P_i}$$

and

$$(39) \quad \gamma \equiv B + \sum \frac{C_i P_i}{\Delta P_i}.$$

Using a Taylor-series expansion argument, a straightforward but notationally very messy lemma shows that the error in the objective function introduced by using the solution of (34)–(36) as an approximation to the solution of (13)–(16) is of second or higher order in $\{\Delta P_i\}$.

Now the problem (34)–(36) is in the form of a classical linear programming budgeting problem. The relevant solution concept is to rank “investment opportunities” by the ratio criterion:

$$(40) \quad \left\{ \frac{\alpha_i}{\beta_i} \right\}$$

and always to favor higher-ranked projects over those of lower rank.

When applied to the Noah’s Ark context, and making use of (37), (38), the ratio criterion (40) is equivalent to following precisely the statement of Theorem 4. This concludes the concise proof of Theorem 4. Q.E.D.

Theorem 4 represents a culmination of the research strategy motivating this paper. It has been shown that a methodology that “feels” like a traditional cost-effectiveness approach can be constructed to deal with the conservation of diversity. The myopic ranking criterion developed here is sufficiently operational to be at least useful in suggesting *what* to look at when determining conservation priorities. At the same time, the formula is rigorously derived from an optimization framework, so that its theoretical foundations are clear.

DISCUSSION

The ranking formula (29) encourages the conservation authorities to focus on four fundamental ingredients when choosing priorities:

- D_i = *distinctiveness* of i = how unique or different is i ;
- U_i = *direct utility* of i = how much we like or value i *per se*;
- ΔP_i = by how much can the *survivability* of i actually be improved;
- C_i = how much does it *cost* to improve the survivability of i .

I am not intending here to argue that it is easy in practice to quantify the above four variables and combine them routinely into the ranking formula (29) that defines R_i . The real world is more than a match for any model. The essential worth of this kind of research is to suggest a framework or way of thinking about biodiversity preservation, and to indicate how it might be backed by a rigorous underlying formulation.

The basic hope is that the formula for R_i could still be used as a rough guide or rule of thumb for deciding conservation priorities even in situations where we cannot know C_i , ΔP_i , D_i , or U_i with any precision. Perhaps one could come away with a sense that when making conservation decisions in the name of preserving diversity, it might seem like a "good idea" at least to consider the four factors D_i , U_i , ΔP_i , and C_i —especially in a policy world so otherwise lacking justifiable guidelines for endangered species protection. One is perhaps further encouraged to think that combining these four ingredients into an overall index R_i , more or less as indicated by (29), also seems like a "good idea" because it is intuitively plausible and backed by a rigorous theory in a special, but not unreasonable, case.

The ultimate hope is that the metaphor of the Noah's Ark Problem, the associated Myopic Ranking Criterion, the underlying Library Model, and the rest of the conceptual apparatus have "staying power" as a useful guide to organized thinking about the economics of biodiversity preservation.

Dept. of Economics, Harvard University, Cambridge, MA 02138, U.S.A.

Manuscript received July, 1996; final revision received June, 1997.

REFERENCES

- BROWN, GARDNER M., AND JOHN H. GOLDSTEIN (1984): A Model for Valuing Endangered Species," *Journal of Environmental Economics and Management*, 11, 303–309.
- CROZIER, ROSS H. (1992): "Genetic Diversity and the Agony of Choice," *Biological Conservation*, 61, 11–15.
- METRICK, ANDREW, AND MARTIN L. WEITZMAN (1994): "Patterns of Behavior in Endangered Species Preservation," *Land Economics*, 72, 1–16.
- REID, W. V., AND K. R. MILLER (1989): *Keeping Options Alive: The Scientific Basis for Conserving Biodiversity*. Washington D.C.: World Resources Institute.

- SOLOW, ANDREW R., STEPHAN POLASKY, AND JAMES M. BROADUS (1993): "On the Measurement of Biological Diversity," *Journal of Environmental Economics and Management*, 24, 60–68.
- (1993): "Searching for Uncertain Benefits and the Conservation of Biological Diversity," *Environmental and Resource Economics*, 3, 171–181.
- WEITZMAN, MARTIN L. (1992): "On Diversity," *Quarterly Journal of Economics*, 107, 363–406.
- (1993): "What to Preserve? An Application of Diversity Theory to Crane Conservation," *Quarterly Journal of Economics*, 108, 157–183.