Diversity[*]

Klaus Nehring
Department of Economics, University of California at Davis
Davis, CA 95616, U.S.A.

and

Clemens Puppe
Department of Economics, University of Karlsruhe
D – 76128 Karlsruhe, Germany

forthcoming in:

The Handbook of Rational and Social Choice
edited by P. Anand, P. Pattanaik and C. Puppe
Oxford University Press

---

[*] We thank Stefan Baumgärtner and Nicolas Gravel for helpful comments and suggestions.

1 Introduction

How much species diversity is lost in the Brazilian rain forest every year? Is France culturally more diverse than Great Britain? Is the range of car models offered by BMW more or less diverse than that of Mercedes-Benz? And more generally: What is diversity, and how can it be measured?

This chapter critically reviews the recent attempts in the economic literature to answer this question. As indicated, the interest in a workable theory of diversity and its measurement stems from a variety of different disciplines. From an economic perspective, one of the most urgent global problems is the quantification of the benefits of ecosystem services and the construction of society's preferences over different conservation policies. In this context, biodiversity is a central concept that still needs to be understood and appropriately formalized. In welfare economics, it has been argued that the range of different life-styles available to a person is an important determinant of this person's well-being (see, e.g., Dowding and van Hees, 2007). Again, the question arises how this range, can be quantified. Finally, the definition and measurement of product diversity in models of monopolistic competition and product differentiation is an important and largely unresolved issue since Dixit and Stiglitz's (1977) seminal contribution.

The central task of a theory of diversity is to properly account for the similarities and dissimilarities between objects. In the following, we present some basic approaches to this problem.[1]

2 Measures based on Dissimilarity Metrics

A natural starting point to think about diversity is based on the intuitive inverse relationship between diversity and similarity: the more dissimilar objects are among each other, the more diverse is their totality. Clearly, this approach is fruitful only to the extent to which our intuitions about (dis)similarity are more easily accessible than those about diversity. In the following, we distinguish the different concrete proposals according to the nature of the underlying dissimilarity relation: whether it is understood as a binary, ternary, or quaternary relation, and whether it is used as a cardinal or only ordinal concept.

2.1 Ordinal Notions of Similarity and Dissimilarity

Throughout, let X denote a finite universe of objects. As indicated in the introduction, the elements of X can be as diverse objects as biological species, ecosystems, life-styles, brands of products, the flowers in the garden of your neighbor, etc. The simplest notion of similarity among the objects in X is the dichotomous distinction according to which two elements are either similar or not, with no intermediate possibilities. Note that in almost all interesting cases such binary similarity relations will not be transitive. Pattanaik and Xu (2000) have used this simple notion of similarity in order to define a ranking of sets in terms of diversity, as follows. A *similarity-based partition* of a set $S \subseteq X$ is a partition $\{A_1, ..., A_m\}$ of S such that, for each partition element A_i, all elements in A_i are similar to each other. Clearly, similarity-based partitions

---

[1] For recent alternative overviews, see Baumgärtner (2006) and Gravel (2007).

thus defined are in general not unique. As a simple example, consider the universe $X = \{x, y, z\}$ and suppose that x and y, as well as y and z are similar, but x and z are not similar. The singleton-partition (i.e. here: $\{\{x\}, \{y\}, \{z\}\}$) always qualifies as a similarity-based partition. In addition, there are the following two similarity-based partitions in the present example, namely $\{\{x, y\}, \{z\}\}$ and $\{\{x\}, \{y, z\}\}$. Pattanaik and Xu (2000) propose to take the minimal cardinality of a similarity-based partition of a set as an ordinal measure of its diversity.

Evidently, the ranking proposed (and axiomatized) by Pattanaik and Xu (2000) is very parsimonious in its informational requirements. Inevitably, this leads to limitations in its applicability, since differential degrees of similarity often appear to have a significant effect on the entailed diversity. To enrich the informational basis while sticking to the ordinal framework, Bervoets and Gravel (2007) have considered a quaternary similarity relation that specifies which *pairs* of objects are comparably more dissimilar to each other than other pairs of objects.[2] Bervoets and Gravel (2007) axiomatize the "maxi-max" criterion according to which a set is more diverse than another if its two most dissimilar elements are more dissimilar than those of the other set.[3] One evident problem with this approach (and the ordinal framework, more generally) is that it cannot account for trade-offs between the number and the magnitude of binary dissimilarities. Intuitively, it is by no means evident that a set consisting of two maximally dissimilar elements is necessarily more diverse than a set of many elements all of which are pairwise less dissimilar. In order to properly account for such trade-offs one needs cardinal dissimilarity information.

2.2 Cardinal Dissimilarity Metrics

In a seminal contribution, Weitzman (1992) has proposed to measure diversity based on a cardinal dissimilarity metric, as follows. Let $d(x, y)$ denote the dissimilarity between x and y, and define the *marginal diversity* of an element x at a given set S by

$$v(S \cup \{x\}) - v(S) = \min_{y \in S} d(x, y). \tag{2.1}$$

Given any valuation of singletons (i.e. sets containing only one element), and given any ordering of the elements $x_1, ..., x_m$, formula (2.1) recursively yields a diversity value $v(S)$ for the set $S = \{x_1, ..., x_m\}$.[4] The problem is that the resulting value in general depends on the ordering of the elements. Weitzman (1992) observes this, and shows that (2.1) can be used to assign a unique diversity value $v(S)$ to each set S if and only if d is an *ultrametric*, i.e. a metric with the additional property that the two greatest distances between three points are always equal.[5] To overcome the restrictiveness of

---

[2] Denoting the quaternary relation by Q, the interpretation of $(x, y)Q(z, w)$ is thus that x and y are more dissimilar to each other than z and w. Bossert, Pattanaik and Xu (2003) have also considered relations of this kind and observed that the dichotomous case considered above corresponds to the special case in which Q has exactly two equivalence classes.

[3] The maximal distance between any two elements is often called the *diameter* of a set. The ranking of sets according to their diameter has been also proposed in the related literature on freedom of choice by Rosenbaum (2000). In the working paper version, Bervoets and Gravel (2007) also consider a lexicographic refinement of the "maxi-max" criterion.

[4] Indeed, by (2.1) we have $v(\{x_1, ..., x_k\}) = \min_{i=1,...,k-1} d(x_k, x_i) + v(\{x_1, ..., x_{k-1}\})$ for all $k = 2, ..., m$. Thus, given the ordering of elements, $v(\{x_1, ..., x_m\})$ can be recursively determined from the dissimilarity metric and the value $v(\{x_1\})$.

[5] Such metrics arise naturally, e.g. in evolutionary trees as shown by Weitzman (1992), see Section 3.2 below for further discussion.

formula (2.1), Weitzman (1992) has also proposed a more general recursion formula. However, the entailed diversity evaluation of sets has the counterintuitive property that the marginal diversity of an object can *increase* with the set to which it is added (see Section 3.1 below for further discussion). The ordinal ranking corresponding to Weitzman's general recursion formula has been axiomatically characterized by Bossert, Pattanaik and Xu (2003).

The fact that the validity of formula (2.1) is restricted to ultrametrics reveals a fundamental difficulty in the general program to construct appropriate diversity measures from binary dissimilarity information (see van Hees, 2004, for further elaboration of this point). There do not seem to exist simple escape routes. For instance, ranking sets according to the average dissimilarity, i.e. $v(S) = \sum_{\{x,y\} \subseteq S} d(x,y)/\#S$, is clearly inappropriate due to the discontinuity when points get closer to each other and merge in the limit; other measures based on the sum of the dissimilarities have similar problems. We therefore turn to an alternative approach that has been suggested in the literature.

3 The Multi-Attribute Model of Diversity

In a series of papers (Nehring and Puppe, 2002, 2003, 2004a, 2004b), we have developed a *multi-attribute approach* to valuing and measuring diversity. Its basic idea is to think of the diversity of a set as derived from the number and weight of the different *attributes* possessed by its elements.[6] Due to its generality, the multi-attribute approach allows one to integrate and compare different proposals of how diversity is based on binary dissimilarity information, and to ask questions such as "when, in general, can diversity be determined by binary information?"

3.1 The Basic Framework

As a simple example in the context of biodiversity, consider a universe X consisting of three distinct species: whales (wh), rhinoceroses (rh) and sharks (sh). Intuitively, judgements on the diversity of different subsets of these species will be based on their possessing different *features.* For instance, whales and rhinos possess the feature "being a mammal," while sharks possess the feature "being a fish." Let F be the totality of all features deemed relevant in the specific context, and denote by $R \subseteq X \times F$ the "incidence" relation that describes the features possessed by each object, i.e. $(x, f) \in R$ whenever object $x \in X$ possesses feature $f \in F$. A sample of elements of R in our example is thus (wh, f_{mam}), (rh, f_{mam}), and (sh, f_{fish}), where f_{mam} and f_{fish} denote the features "being a mammal" and "being a fish," respectively. For each relevant feature $f \in F$, let $\lambda_f \geq 0$ quantify the value of the realization of f. Upon normalization, λ_f can thus be thought of as the relative importance, or *weight* of feature f. The *diversity value* of a set S of species is defined as

$$v(S) := \sum_{f \in F : (x,f) \in R \text{ for some } x \in S} \lambda_f. \tag{3.1}$$

---

[6] Measures of diversity that are based (explicitly or implicitly) on the general idea of counting attributes ("features," "characteristics") have been proposed frequently in the literature, see among others, Vane-Wright, Humphries and Williams (1991), Faith (1992, 1994), Solow, Polasky and Broadus (1993), Weitzman (1998), and the volumes edited by Gaston (1996) and Polasky (2001).

Hence, the diversity value of a set of species is given by the total weight of all different features possessed by some species in S. Note especially that each feature occurs at most once in the sum. In particular, each single species contributes to diversity the value of all those features that are not possessed by any already existing species.

The relevant features can be classified according to which sets of objects possess them, as follows. First, there are all idiosyncratic features of the above species, the sets of which we denote by $F_{\{wh\}}$, $F_{\{rh\}}$ and $F_{\{sh\}}$, respectively. Hence, $F_{\{wh\}}$ is the set of all features that are possessed exclusively by whales, and analogously for $F_{\{rh\}}$ and $F_{\{sh\}}$. For instance, sharks being the only fish in this example, $F_{\{sh\}}$ contains the feature "being a fish." On the other hand, there will typically exist features jointly possessed by several objects. For any subset $A \subseteq X$ denote by F_A the set of features that are possessed by *exactly* the objects in A; thus, each feature in F_A is possessed by all elements of A and not possessed by any element of $X \setminus A$. For instance, whales and rhinos being the only mammals in the example, the feature "being a mammal" belongs to the set $F_{\{wh,rh\}}$. With this notation, (3.1) can be rewritten as

$$v(S) := \sum_{A\cap S\neq\emptyset} \sum_{f\in F_A} \lambda_f. \tag{3.1$'$}$$

Intuitively, any feature shared by several objects corresponds to a similarity between these objects. For instance, the joint feature "mammal" renders whales and rhinos similar with respect to their taxonomic classification. Suppose, for the moment, that the feature of "being a mammal" is in fact the only non-idiosyncratic feature deemed relevant in our example, and let λ_{mam} denote its weight. In this case, (3.1) or (3.1$'$) yield $v(\{wh, sh\}) = v(\{wh\}) + v(\{sh\})$, i.e. the diversity value of whale and shark species together equals the sum of the value of each species taken separately. On the other hand, since $v(\{wh, rh\}) = v(\{wh\})+v(\{rh\})-\lambda_{mam}$, the diversity value of whale and rhino species together is *less* than the sum of the corresponding individual values by the weight of the common feature "mammal." This captures the central intuition that the diversity of a set is reduced by similarities between its elements.

It is useful to suppress explicit reference to the underlying description F of relevant features by identifying features *extensionally.* Specifically, for each subset $A \subseteq X$ denote by $\lambda_A := \sum_{f\in F_A} \lambda_f$ the total weight of all features with extension A, with the convention that $\lambda_A = 0$ whenever $F_A = \emptyset$. With this notation, (3.1$'$) can be further rewritten as

$$v(S) = \sum_{A\cap S\neq\emptyset} \lambda_A. \tag{3.1$''$}$$

The totality of all features $f \in F_A$ will be identified with their extension A, and we will refer to the subset A as a particular *attribute.* Hence, a set A viewed as an attribute corresponds to the family of all features possessed by exactly the elements of A. For instance the attribute $\{wh\}$ corresponds to the conjunction of all idiosyncratic features of whales ("being a whale"), whereas the attribute $\{wh, rh\}$ corresponds to "being a mammal."[7] The function λ that assigns to each attribute A its weight λ_A, i.e. the total weight of all features with extension A, is referred to as the *attribute weighting*

---

[7] Subsets of X thus take on a double role as sets to be evaluated in terms of diversity on the one hand, and as weighted attributes on the other. In order to notationally distinguish these roles we will always denote generic subsets by the symbol "A" whenever they are viewed as attributes, and by the symbol "S" otherwise.

function. The set of *relevant* attributes is given by the set $\Lambda := \{A : \lambda_A \neq 0\}$. An element $x \in X$ *possesses* the attribute A if $x \in A$, i.e. if x possesses one, and therefore all, features in F_A. Furthermore, say that an attribute A is *realized* by the set S if it is possessed by at least one element of S, i.e. if $A \cap S \neq \emptyset$. According to (3.1″), the diversity value $v(S)$ is thus the total weight of all attributes realized by S.

A function v of the form (3.1″) with $\lambda_A \geq 0$ for all A is called a *diversity function*, and we will always assume the normalization $v(\emptyset) = 0$. Clearly, any given attribute weighting function $\lambda \geq 0$ determines a particular diversity function via formula (3.1″). Conversely, any given diversity function v *uniquely* determines the corresponding collection λ_A of attribute weights via "conjugate Moebius inversion."[8] In particular, any given diversity function v unambiguously determines the corresponding family Λ of relevant attributes. This basic fact allows one to describe properties of a diversity function in terms of corresponding properties of the associated attribute weighting function.

An essential property of a diversity function is that the marginal value of an element x *decreases* in the size of existing objects; formally, for all S, T and x

$$S \subseteq T \Rightarrow v(S \cup \{x\}) - v(S) \geq v(T \cup \{x\}) - v(T). \tag{3.2}$$

Indeed, using (3.1″), one easily verifies that

$$v(S \cup \{x\}) - v(S) = \sum_{A \ni x, A \cap S = \emptyset} \lambda_A,$$

which is decreasing in S due to the non-negativity of λ. Property (3.2), known as *submodularity*, is a very natural property of diversity; it captures the fundamental intuition that it becomes the harder for an object to add to the diversity of a set the larger that set already is.[9]

Any diversity function naturally induces a notion of pairwise dissimilarity between species. Specifically, define the *dissimilarity from x to y* by

$$d(x, y) := v(\{x, y\}) - v(\{y\}). \tag{3.3}$$

The dissimilarity $d(x, y)$ from x to y is thus simply the marginal diversity of x in a situation in which y is the only other existing object. Using (3.1″) one easily verifies that

$$d(x, y) = \sum_{A \ni x, A \not\ni y} \lambda_A,$$

that is, $d(x, y)$ equals the weight of all attributes possessed by x but not by y. Note that, in general, d need not be symmetric, and thus fails to be a proper metric; it does, however, always satisfy the triangle inequality. The function d is symmetric if and only if $v(\{x\}) = v(\{y\})$ for all $x, y \in X$, i.e. if and only if single objects have identical diversity value.

A decision-theoretic foundation of our notion of diversity can be given along the lines developed by Nehring (1999a). Specifically, it can be shown that a von-Neumann-Morgenstern utility function v derived from ordinal expected utility preferences over

---

[8] Specifically, one can show that if a function v satisfies (3.1″) for all S, then the attribute weights are (uniquely) determined by $\lambda_A = \sum_{S \subseteq A} (-1)^{\#(A \setminus S)+1} \cdot v(X \setminus S)$, see Nehring and Puppe (2002, Fact 2.1).

[9] A somewhat stronger property, called *total* submodularity, in fact characterizes diversity functions, see Nehring and Puppe (2002, Fact 2.2).

distributions of sets of objects is a diversity function, i.e. admits a non-negative weighting function λ satisfying $(3.1'')$, if and only if the underlying preferences satisfy the following axiom of "indirect stochastic dominance:" a distribution of sets p is (weakly) preferred to another distribution q whenever, for all attributes A, the probability of realization of A is larger under p than under q (see Nehring, 1999a, and Nehring and Puppe, 2002, for details). In this context, distributions of sets of objects can be interpreted in two ways: either as the uncertain consequences of conservation policies specifying (subjective) survival probabilities for sets of objects, or as describing (objective) frequencies of sets of existing objects, e.g. as the result of a sampling process. In terms of interpretation, different preferences over probabilistic lotteries describe different *valuations* of diversity (or, equivalently, of the realization of attributes). By contrast, different rankings of frequency distributions correspond to different ways of *measuring* diversity. The multi-attribute approach is thus capable to incorporate either the valuation and the measurement aspect of diversity.[10]

3.2 Diversity as Aggregate Dissimilarity

In practical applications, one will have to construct the diversity function from primitive data. One possibility is, of course, to first determine appropriate attribute weights and to compute the diversity function according to $(3.1'')$. Determining attribute weights is a complex task, however, since there are as many potential attributes as there are non-empty *subsets* of objects, i.e. $2^n - 1$ when there are n objects. An appealing alternative is to try to derive the diversity of a set from the pairwise dissimilarities between its elements. This is a much simpler task since, with n objects, there are at most $n \cdot (n-1)$ non-zero dissimilarities. The multi-attribute approach makes it possible to determine precisely when the diversity of a set can be derived from the pairwise dissimilarities between its elements. The central concept is that of a "model of diversity."

A non-empty family of attributes $\mathcal{A} \subseteq 2^X \backslash \{\emptyset\}$ is referred to as a *model (of diversity).* A diversity function v is *compatible* with the model $\mathcal{A}$ if the corresponding set Λ of relevant attributes is contained in $\mathcal{A}$, i.e. if $\Lambda \subseteq \mathcal{A}$. A model thus represents a *qualitative* a priori restriction, namely that no attributes outside $\mathcal{A}$ can have strictly positive weight. For instance, in a biological context, an example of such an a priori restriction would be the requirement that all relevant attributes are biological taxa, such as "being a vertebrate," "being a mammal," etc. This requirement leads to an especially simple functional form of any compatible diversity function, as follows. Say that a model $\mathcal{A}$ is *hierarchical* if, for all $A, B \in \mathcal{A}$ with $A \cap B \neq \emptyset$, either $A \subseteq B$ or $B \subseteq A$. In Nehring and Puppe (2002) it is shown that a diversity function v is compatible with a hierarchical model if and only if, for all S,

$$v(S \cup \{x\}) - v(S) = \min_{y \in S} d(x, y),$$

where d is defined from v via (3.3). This is precisely Weitzman's recursion formula (2.1) with the only difference that no symmetry of d is required here. Thus, Weitzman's original intuition turns out to be correct exactly in the hierarchical case.[11]

---

[10] For an argument that the measurement of diversity presupposes some form of value judgment, see Baumgärtner (2007).

[11] Another example of a hierarchical model emerges by taking the "clades" in the evolutionary tree, i.e. for any species x the set consisting of x and all its descendants, as the set of relevant attributes.

A more general model that still allows one to determine the diversity of arbitrary sets from the binary dissimilarities between its elements is the *line model.* Specifically, suppose that the universe of objects X can be ordered from left to right in such a way that all relevant attributes are connected subsets, i.e. intervals. This structure emerges, for instance, in the above example once one includes the non-taxonomic attribute "ocean-living" possessed by whales and sharks (see Figure 1).

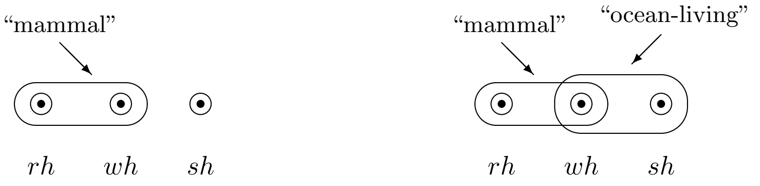

Figure 1: Hierarchical versus linear organization of attributes

A diversity function v is compatible with this *line model* if and only if, for all sets $S = \{x_1, ..., x_m\}$ with $x_1 \leq x_2 \leq ... \leq x_m$,

$$v(S) = v(\{x_1\}) + \sum_{i=2}^{m} d(x_i, x_{i-1}) \tag{3.4}$$

(see Nehring and Puppe, 2002, Theorem 3.2).

When, in general, is diversity determined by binary information alone? Say that a model $\mathcal{A}$ is *monotone in dissimilarity* if, for any compatible diversity function v and any S, the diversity $v(S)$ is uniquely determined by the value of all single elements in S and the pairwise dissimilarities within S, and if, moreover, the diversity $v(S)$ is a monotone function of these dissimilarities. Furthermore, say that a model $\mathcal{A}$ is *acyclic* if for no $m \geq 3$ there exist elements $x_1, ..., x_m$ and attributes $A_1, ..., A_m \in \mathcal{A}$ such that, for all $i = 1, ..., m-1$ $A_i \cap \{x_1, ..., x_m\} = \{x_i, x_{i+1}\}$, and $A_m \cap \{x_1, ..., x_m\} = \{x_m, x_1\}$. Thus, for instance in the case $m = 3$, acyclicity requires that there be no triple of elements such that each pair of them possesses an attribute that is not possessed by the third element. A main result of Nehring and Puppe (2002) establishes that a model of diversity is monotone in dissimilarity if and only if it is acyclic.[12]

An important example of a non-acyclic model is the *hypercube model* which takes the set of all binary sequences of length K ("the K-dimensional hypercube") as the universe of objects and assumes all relevant attributes to be subcubes (i.e. subsets forming a cube of dimension $k \leq K$).[13] The hypercube model is clearly not acyclic (see Nehring and Puppe, 2002, Section 3.3). To illustrate the possible violations of monotonicity in dissimilarity in the hypercube model, consider the following five points in the 4-hypercube: $a = (0,0,0,0)$, $b = (0,0,1,1)$, $c = (1,0,1,0)$, $x_1 = (0,1,1,0)$ and $x_2 = (1,0,0,1)$ (see Figure 2). If all subcubes of the same dimension have the

---

For a critique of the "cladistic model" and an alternative proposal, the "phylogenetic tree model," see Nehring and Puppe (2004a).

[12] The necessity of acyclicity hinges on a weak regularity requirement, see Nehring and Puppe (2002, Section 6).

[13] The hypercube model seems to be particularly appropriate in the context of sociological diversity. In this context, individuals are frequently classified according to binary characteristics such as "male vs. female," "resident vs. non-resident" etc.

same (positive) weight, then the dissimilarity $d(y, z)$ is uniquely determined by the Hamming distance between y and z.[14] Now consider the sets $S_1 = \{a, b, c, x_1\}$ and $S_2 = \{a, b, c, x_2\}$. The two sets are metrically isomorphic since any element in either set has Hamming distance 2 from any other element in the same set. Nevertheless S_1 is unambiguously more diverse since S_2 is entirely contained in the 3-dimensional subcube spanned by all elements with a "0" in the second coordinate (the white front cube in Figure 2). By contrast, S_1 always gives a choice between "0" and "1" in each coordinate.

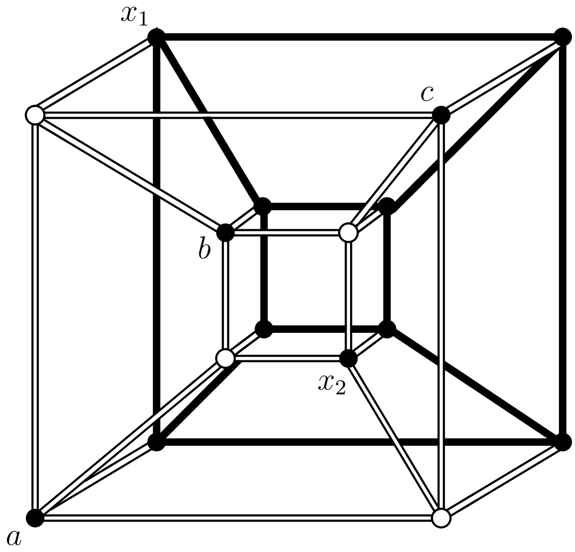

Figure 2: Two metrically isomorphic subsets of the 4*-hypercube*

3.3 On the Application of Diversity Theory

In the context of biodiversity a key issue is the choice of an appropriate conservation policy such as investment in conservation sites, restrictions of land development, anti-poaching measures, or the reduction of carbondioxid emission. This can be modeled along the following lines. A policy determines at each point of time a probability distribution over sets of existing species and consumption. Formally, a policy p can be thought of a sequence $p = (p^t)_{t \geq 0}$, where each p^t is a probability distribution on $2^X \times \mathbf{R}_+^N$ with $p^t(S^t, c^t)$ as the probability that at time t the set S^t is the set of existing species and c^t is the consumption vector. Denoting by P the set of feasible policies, society's problem can thus be written as

$$\max_{p \in P} \int_0^\infty e^{-\delta t} \cdot E_{p^t}[v(S^t) + u(c^t)]dt, \tag{3.5}$$

where δ denotes the discount rate and E_p the expectation with respect to p. The objective function in (3.5) is composed of utility from aggregate consumption $u(c^t)$ and

---

[14] By definition, the Hamming distance between two points in the hypercube is given by the number of coordinates in which they differ.

the existence value $v(S^t)$ from the set S^t of surviving species; its additively separable form is assumed here for simplicity.

Diversity theory tries to help us determine the intrinsic value we put on the survival of different species which is represented by the function v. The probabilities p^t reflect societies expectations about the consequences of its actions; these, in turn, reflect our knowledge of economic and ecological processes. For instance, the role of keystone species that are crucial for the survival of an entire ecosystem will be captured in the relevant probability distribution. Thus, the value derived from the presence of such species *qua* keystone species enters as an indirect rather than intrinsic utility.[15]

As a simple example, consider two species y and z each of which can be saved forever (at the same cost); moreover, suppose that it is not possible to save both of them. Which one should society choose to save? Assuming constant consumption *ceteris paribus*, the utility gain at t from saving species x, given that otherwise the set S^t of species survives, is

$$v(S^t \cup \{x\}) - v(S^t) = \sum_{A \ni x, A \cap S^t = \emptyset} \lambda_A.$$

Denote by $Q^t(x) := \sum_{A \ni x} \lambda_A \cdot prob(A \cap S^t = \emptyset)$ the expected marginal value at t of saving x, which is given by the sum of the weights of all attributes possessed by x multiplied by the probability that x is the unique species possessing them. The expected present value of the utility gain from saving x is given by

$$\int_0^\infty e^{-\delta t} \cdot Q^t(x) dt.$$

For concreteness, let y be one of the few species of rhinoceroses, and z a unique endemic species which currently has a sizeable number of fairly distant relatives. In view of the fact that all rhino species are currently endangered, this leads to the following trade-off between maximizing diversity in the short-run and in the long-run. Saving the endemic species z yields a significant short-run benefit, while the expected benefit from safeguarding the last rhino species would be very high. This suggests the following qualitative behavior of the streams of intertemporal benefits accruing from the two policies:

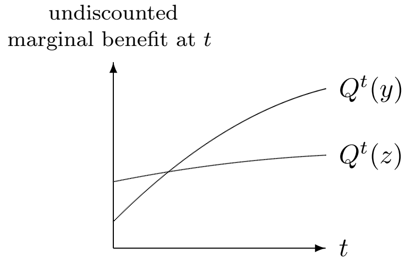

Figure 3: Streams of expected marginal benefits

---

[15] Alternatively, the multi-attribute framework can also be interpreted in terms of *option value*, as explained in Nehring and Puppe (2002, p.1168). As a result, measures of biodiversity based on that notion, such as the one proposed in Polasky, Solow and Broadus (1993), also fit into the framework of the multi-attribute model.

The strong increase in the expected marginal value of saving y stems from the fact that, due to the limited current number of rhinos, the extinction probability of their unique attributes becomes high as t grows. Clearly, the rhino species y should be saved if the discount rate is low enough; otherwise, z should be saved. The decision thus depends on three factors: the discount rate, the value of the relevant attributes at stake, and the probability of the survival of close relatives over time.

4 Abstract Convexity and the Geometry of Similarity

4.1 Convex Models described by Structural Similarity Relations

A key issue in applications of diversity theory is the danger of combinatorial explosion, since the number of conceivable attributes, and hence the upper bound on the number of independent value assessments, grows exponentially in the number of objects. Nehring (1999b) proposes a general methodology of taming this combinatorial explosion refining the idea of a model as a family of (potentially relevant) attributes $\mathcal{A} \subseteq 2^X \setminus \{\emptyset\}$ introduced in Section 3.2.

The key idea is to assume that the family of potentially relevant attributes is *patterned* in an appropriate way. Such patterning is important for two related reasons. First, excluding an isolated attribute rather than a patterned set of attributes typically does not correspond to an interpretable restriction on preferences.[16] Second, an isolated exclusion of an attribute will not capture a well-defined structural feature of the situation to be modeled.

Nehring (1999b) argues that an appropriate notion of pattern is given by that of an "abstract convex structure" in the sense of abstract convexity theory.[17] To motivate it, consider the case of objects described in terms of an ordered, "one-dimensional" characteristic such as mass for species or latitude for habitats. Here, the order structure motivates a selection of attributes of the form "weighs no more than 20 grams," "weighs at least 1 ton," "weighs between 3 and 5 kilograms," that is: of *intervals* of real numbers. This selection defines the "line model" introduced in Section 3.2; it rules out, e.g., the conceivable attribute "weighs an odd number of grams".

Any family of relevant attributes $\mathcal{A}$ induces a natural ternary *structural similarity relation* $T_{\mathcal{A}}$ on objects as follows: y is *at least as similar to* z *than* x *is to* z if y shares all relevant attributes with z that x shares with z. In the line model, e.g., in which all attributes are intervals, the weight "5 kilograms" shares all attributes with the weight "10 kilograms" that the weight "1 kilogram" does; by contrast, the weight "1 ton" does not share all attributes common to "10 kilograms" and "1 kilogram." Likewise, in a hierarchical model in which the set of relevant attributes of species is given by

---

[16] In view of conjugate Moebius inversion (see Section 3.1 above), excluding a particular attribute A by imposing the restriction "$\lambda_A = 0$" is equivalent to a linear equality on v involving $2^{\#(X \setminus A)-1}$ terms which will lack a natural interpretation unless $\#(X \setminus A)$ is very small. In Nehring and Puppe (2004b) it is shown more specifically that this restriction can be viewed as a restriction on a $\#(X \setminus A)$-th order partial derivative (more properly: $\#(X \setminus A)$-th order partial difference) of the diversity function.

[17] Abstract convexity theory is a little known field of combinatorial mathematics whose neighboring fields include lattice and order theory, graph theory, and axiomatic geometry. It is surveyed in the rich monograph by van de Vel (1993).

biological taxonomy, a chimpanzee is at least as similar to human than a pig is, since the chimpanzee shares all taxonomic attributes with a human that a pig does.

A family of attributes can now be defined as "patterned" if it is determined by its similarity geometry $T_{\mathcal{A}}$. To do so, one can associate with any ternary relation T on X (i.e. any $T \subseteq X \times X \times X$) an associated family $\mathcal{A}_T$ by stipulating that $A \in \mathcal{A}_T$ if, for any $(x, y, z) \in T$, $\{x, z\} \subseteq A$ implies $y \in A$. A family of attributes $\mathcal{A}_T$ derived from some T satisfies three properties: Boundedness ($\emptyset, X \in \mathcal{A}$), Intersection Closure ($A, B \in \mathcal{A}$ implies $A \cap B \in \mathcal{A}$) and Two-Arity to be defined momentarily. These three properties define a *convex model.* The second is the most important of the three. Translated into the language of attributes, it says that an arbitrary *conjunction* of relevant attributes is a relevant attribute. For example, if "mammal" and "ocean-living" are relevant attribute, so is the conjoint attribute "is a mammal and lives in the ocean."Note that this closure property is much more natural than closure under disjunction; for example, "is a mammal *or* lives in the ocean" is entirely artificial.[18]

The first two properties identify $\mathcal{A}$ as an abstract convex structure in the sense of abstract convexity theory (see van de Vel, 1993). In particular, the first two properties allow one to define, for any $S \subseteq X$ the (abstract) convex hull $co_{\mathcal{A}}(S) := \cap \{A \in \mathcal{A} : A \supseteq S\}$. Two-Arity says that $A \in \mathcal{A}$ whenever A contains, for any $x, y \in A$, their convex hull $co_{\mathcal{A}}(\{x, y\})$. It is easily verified that if the families $\mathcal{A}$ and $\mathcal{B}$ are convex models, so is $\mathcal{A} \cap \mathcal{B}$. It follows that for any family (model) $\mathcal{A} \subseteq 2^X \setminus \emptyset$, there exists a unique smallest superfamily $\mathcal{A}^*$ of $\mathcal{A}$ that is a convex model, the *convexity hull* of $\mathcal{A}$. Nehring (1999b) shows that $\mathcal{A}_{(T_{\mathcal{A}})} = \mathcal{A}^*$ for any $\mathcal{A}$; it follows that $\mathcal{A}$ is a convex model if and only if $\mathcal{A} = \mathcal{A}_{(T_{\mathcal{A}})}$. Thus convex models are exactly the models that are characterized by their associated qualitative similarity relation $T_{\mathcal{A}}$.

Structural similarity relations are characterized by transitivity and symmetry properties; symmetry in particular means that if y is at least as similar to z than x is to z, then y must also be at least as similar to x than z is to x. In view of these properties, structural similarity can be interpreted geometrically as betweenness ("y lies between x and z"). For example, structural similarity in the line model is evidently nothing but the canonical notion of betweenness on a line: y lies between x and z if and only if $x \geq y \geq z$ or $x \leq y \leq z$. A structural similarity relation can therefore be viewed as describing the *similarity geometry* of the space of objects. This endows a convex model with the desired qualitative interpretation.

4.2 Structural Similarity Revealed

Besides this direct conceptual significance, structural similarity relations are useful because they directly relate the structure of the support of λ to that of the diversity function itself. In the following, denote by $d(x, S) := v(S \cup \{x\}) - v(S)$ the marginal value of x at S (the "distinctiveness" of x from S). Say that x is *revealed at least as similar to* z than y – formally, $(x, y, z) \in T_v$ – if $d(x, \{y\}) = d(x, \{y, z\})$. To understand the definition, note that

$$d(x, \{y\}) - d(x, \{y, z\}) = \sum_{A: x \in A, y \notin A} \lambda_A - \sum_{A: x \in A, y \notin A, z \notin A} \lambda_A = \sum_{A: x \in A, y \notin A, z \in A} \lambda_A .$$

---

[18] In a related vein, the philosopher Gärdenfors has argued in a series of papers (see, e.g. Gärdenfors, 1990) that legitimate inductive inference needs to be based on convex predicates.

By non-negativity of λ, one always has $d(x, \{y\}) \geq d(x, \{y, z\})$; moreover, $d(x, \{y\}) > d(x, \{y, z\})$ if and only if a single term on the right-hand side is positive, i.e. if there exists an attribute $A \in \Lambda$ that is common to x and z but not possessed by y. But this simply says that, for any diversity function v, the revealed similarity T_v is identical to the similarity associated with the family of relevant attributes T_Λ,

$$T_v = T_\Lambda.$$

This result has the following two important corollaries. The first characterizes compatibility with a convex model: for any convex model $\mathcal{A}$ and any diversity function with corresponding set Λ of relevant attributes,

$$\Lambda \subseteq \mathcal{A} \Leftrightarrow T_v \supseteq T_{\mathcal{A}}. \tag{4.1}$$

The second corollary shows that the set of relevant attributes is revealed from T_v "up to abstract convexification:"for any diversity function v, $\Lambda^* = \mathcal{A}_{(T_v)}$.

The equivalence (4.1) is as powerful as it is simple, since it amounts to a universal characterization result for arbitrary convex models. For example, noting that for diversity functions, $(x, y, z) \in T_v$ is equivalent to the statement that $d(x, \{y\}) = d(x, S)$ for any S containing y, it allows one to deduce the line equation (3.4) and the hierarchy recursion (2.1) straightforwardly.

4.3 Application to Multi-Dimensional Settings

An important application of (4.1) is to the characterization of multi-dimensional models in which X is the Cartesian product of component spaces, $X = \Pi_k X_k$; an example is the hypercube introduced in Section 3.2. In the context of biodiversity, multi-dimensional models arise naturally if diversity is conceptualized in functional, morphological[19] or genetic rather than in, or in addition to, phylogenetic terms. In multi-dimensional settings, it is natural to require that any relevant attribute share this product structure as well, i.e. that $\Lambda \subseteq \mathcal{A}_{sep}$, where $\mathcal{A}_{sep}$ is the set of all $A \subseteq X$ of the form $A = \Pi_k A_k$. Diversity functions with this property are called *separable.* Since $\mathcal{A}_{sep}$ is easily seen to be a convex model, the equivalence (4.1) can be applied to yield a straightforward characterization of separability that allows one to check whether the restrictions on diversity values/preferences imposed by this mathematically convenient assumption are in fact reasonable. Indeed, $(x, y, z) \in T_{\mathcal{A}^{sep}}$ if and only if, for all k, $y^k \in \{x^k, z^k\}$. Thus separability amounts to the requirement that $d(x, \{y\}) = d(x, \{y, z\})$ for all x, y, z such that, for all $k \in K$, $x^k = z^k \Rightarrow y^k = x^k = z^k$.

Note the substantial gains in parsimony: while $X = \Pi_k X_k$ allows for $2^{\Pi_k \# X_k} - 1$ conceivable attributes, $\#\mathcal{A}_{sep} = \Pi_k \left(2^{\# X_k} - 1\right)$; in the case of the K-dimensional hypercube, for example, $\#\mathcal{A}_{sep} = 3^K$.

Under separability, it is further frequently natural (and mathematically extremely useful) to require *independence* across dimensions, i.e., for any $A = \Pi_k A_k$, $\lambda_A = \Pi_k \lambda^k_{A_k}$ for appropriate marginal attribute weighting functions λ^k; Nehring (1999b) provides simple characterizations of independence in terms of the diversity function and the underlying preference relation. Independence achieves further significant gains

---

[19] The "charisma" of many organisms is closely associated with their anatomy and shape, as in case of the horn of the rhino, the nobility of a crane, the grace of a rose, or the sheer size of a whale.

in parsimony, as now only $\sum_k \left(2^{\#X_k} - 1\right)$ independent attribute weights need to be determined, in the K-dimensional hypercube, for example, $3K$ such weights.

In spite of the obvious importance of multi-dimensional settings, to the best of our knowledge only the pioneering contributions by Solow, Polasky and Broadus (1993) and Solow and Polasky (1994) have tried to model diversity in such settings; we do not survey their work in detail, since their measures are quite special and not well understood analytically.[20]

5 Absolute vs. Relative Conceptions of Diversity

The literature is characterized by two competing intuitive, pre-formal conceptions of diversity that we shall term the "absolute" and the "relative." On the absolute conception, diversity is ontological richness; it has found clear formal expression in the multi-attribute model described in Section 3. On the relative conception, diversity is pure difference, heterogeneity. To illustrate the difference, consider the addition of some object z to the set of objects $\{x, y\}$. On the absolute conception, the diversity can never fall, even if z is a copy of x or very similar to it. By contrast, on the relative conception, the diversity may well fall; indeed, if one keeps adding (near) copies of x, the resulting set would be viewed as nearly homogenous and thus almost minimal in diversity.

In the literature, the relative conception has been articulated via indices defined on probability (i.e. relative frequency) distributions over *types* of objects. In a biological context, these types might be species, and the probability mass of a species may be given by the physical mass of all organisms of that species as a fraction of the total mass; in a social context, types might be defined by socio-economic characteristics, and the probability mass of a type be given by the relative frequency of individuals with the corresponding characteristics.

Formally, let $\Delta(X)$ denote the set of all probability distributions on X, with $p \in \Delta(X)$ written as $(p_x)_{x\in X}$, where $p_x \geq 0$ for all x and $\sum_{x\in X} p_x = 1$. Thus, p_x is the fraction of the population of type $x \in X$. The support of p is the set of types with positive mass, $\operatorname{supp} p = \{x \in X : p_x > 0\}$. A *heterogeneity index* is a function $h : \Delta(X) \rightarrow \mathbf{R}$.[21] It is natural to require that h take values between 1 and $\#X$, as this allows an interpretation of "effective number of different types" (cf. Hill, 1973). As developed in the literature, a heterogenity index is understood to rely on the frequency distribution over different types as the *only* relevant information; heterogeneity indices are thus required to be *symmetric,* i.e. invariant under arbitrary permutations of the p-vector. This reflects the implicit assumption that all individuals are either exact copies or just different (by belonging to different types); all non-trivial similarity information among types is ruled out.

To be interpretable as a heterogeneity index, h must rank more "even" distribu-

---

[20] The former paper represents objects as points in a finite-dimensional Euclidean space, and restricts relevant attributes to be balls in this space. The latter provides a lower bound on diversity values of arbitrary sets given the diversity values of sets with at most two elements; it also proposes taking these lower bounds as a possibly useful diversity measure based on distance information in its own right with an interesting statistical interpretation. It seems doubtful that this measure will be ordinarily a diversity function, and thus that it will admit a multi-attribute interpretation.

[21] We use this non-standard terminology to clearly distinguish heterogeneity indices from diversity functions in terms of both their formal structure and their conceptual motivation.

tions higher than less even ones; formally *Preference for Evenness* is captured by the requirement that h be quasi-concave. Note that Symmetry and Preference for Evenness imply that the uniform distribution $(\frac{1}{n}, ..., \frac{1}{n})$ has maximal heterogeneity.

A particular heterogeneity index h is characterized in particular how it trades off the "richness" and the "evenness" of distributions. Roughly, richness measures how many different entities there are (with any non-zero frequency), while evenness measures how frequently they are realized. For instance, comparing the distributions $p = (0.6, 0.3, 0.1)$ and $q = (0.5, 0.5, 0)$, intuitively the former is richer while the latter is more even.

The most commonly used heterogeneity indices belong to the following one-parameter family $\{h_\alpha\}_{\alpha \geq 0}$, in which the parameter $\alpha \geq 0$ describes the trade-off between richness and evenness:

$$h_\alpha (p) = \left(\sum_{x \in X} p_x^\alpha \right)^{\frac{1}{1-\alpha}}.$$

These indices (more properly: their logarithm) are known in the literature as "generalized" or "Renyi" entropies (Renyi, 1961). Like much of the literature, we take these indices to have primarily ordinal meaning; the chosen cardinalization ensures that uniform distributions of the form $(\frac{1}{m}, .., \frac{1}{m}, 0, ...)$ have heterogeneity m. The class of generalized entropy indices $\{h_\alpha\}$ can be cleanly characterized axiomatically; for a nice exposition that draws on a closely related result on inequality measurement by Shorrocks (1984), see Gravel (2007).

A high α implies emphasis on frequent types, and thus a relatively strong weight on evenness over richness. Indeed, in the limit when α grows without bound, one obtains $h_\infty (p) = \frac{1}{\max_{x \in X} p_x}$, i.e. the frequency of the most frequent type determines heterogeneity completely.[22] At the other end of the spectrum ($\alpha = 0$), h_α simply counts the size of the support $\#(\text{supp}\, p)$: here evenness counts for nothing, and richness is everything. Besides the counting index, by far the most important in applications are the parameter values $\alpha = 1$ and $\alpha = 2$.

For $\alpha = 1$, the logarithm of $h_\alpha (p)$ (defined by an appropriate limit operation) is the Shannon-Wiener entropy, $\log_2 h_1 (p) = - \sum_{x \in X} p_x \log_2 p_x$. An intuitive connection to a notion of diversity as disorder comes from its origin in coding theory, where it describes the minimum average number of bits needed to code without redundancy a randomly drawn member from the population.

For $\alpha = 2$, $h_2 (p) = (\sum_{x \in X} p_x^2)^{-1}$ is an ordinal transform of the Simpson index (Simpson, 1949) in the biological literature. Again, an intuitive link to some notion of heterogeneity can be established by noting that $\sum_{x \in X} p_x^2$ is the probability that two randomly and independently drawn elements from the population belong to the same class.

In contrast to the popularity of generalized entropy indices, their conceptual foundations remain to be clarified. We note three issues in particular. First, an important conceptual gap in the existing literature is the lack of a substantive interpretation of the parameter α: what does the parameter α represent? On what grounds should a diversity assessor choose one value of α rather than another? Could α represent a feature of the world; if so, what could that feature be? Alternatively, could α represent

---

[22] The index h_∞ is known as the Berger-Parker index (Berger and Parker, 1970) in the biological literature.

a feature of the assessor, a "taste" for richness versus evenness? Such a preference interpretation may be tempting for economists, especially in view of certain formal similarities to the theory of risk-aversion. Note, however, in the latter the degree of risk-aversion can reasonably (if controversially) be explained, or at least related to, the speed at which the marginal (hedonic) utility decreases with income. The problem with the parameter α is the apparent lack of any such correlate; at least, no such correlate appears to have been suggested in the literature.

Second, the generalized entropy indices rely on a partitional classification of pairs of individuals as either completely identical or completely different. Intermediate degrees of similarity/dissimilarity are ruled out. But these are of evident importance for a relative conception of diversity no less than for an absolute one. In applications, the need to fix a partition introduces a significant degree of arbitrariness into the measurement of heterogeneity.

Third, and perhaps most fundamentally, it is not clear whether the relative conception constitutes a fundamentally different notion of diversity, or whether it is in some way derivable from the absolute conception or, indeed, from a "diversity-free" notion altogether. An example of the latter is Weitzman (2000) model of economically optimal crop variety in which he provides assumptions under which Shannon entropy can serve as a "generalized measure of resistance to extinction."To establish irreducibility, invocations of terms like "surprise" and "disorder" are clearly not enough.[23] While they may serve to visualize notions of (generalized) entropy, they do not establish their appropriateness of these as measures of diversity. Hill (1973, p.428), for example, emphatically asserts that "the information-theoretic analogy is not illuminating."

In the remainder of this chapter, we sketch one way to make sense of relative diversity as derived from absolute diversity by "sampling."The sampling could represent a future evolution/survival process that selects a subset of the given set of individuals. Alternatively, the sampling may capture the diversity experienced by an embodied diversity consumer whose physical or mental eye is constrained by the limited capacity to take in and absorb the existing range of objects. For concreteness, think, for example, of a tourist on an ecotrip. Under both interpretations, the addition of a common organism may hinder the likelihood of survival (respectively of observation) of a less common one, in line with the Preference for Evenness intuition that is characteristic of the relative conception. To come up with a determinate and simple functional form, we will assume a very stylized sampling process with fixed sample size, independent draws and replacement. By building on the multi-attribute model described in Section 3, the resulting family of indices allows one to capture non-trivial similarities in a very general manner. Furthermore, the sample size can serve as an interpretable parameter determining the richness-evenness trade-off. The exposition will be heuristic and hopes to stimulate further research in this important grey area of diversity theory.

Think of individual entities ("individuals") $y \in Y$ as described by their type $x \in X$ and a numeric label $i \in \mathbf{N}$. Thus the domain of individuals is given as $Y = X \times \mathbf{N}$. For a given set of individuals $S \subseteq Y$, it is convenient to write $S_x = S \cap (\{x\} \times \mathbf{N})$ for the subset of individuals in type x, and $q_x^S = \#S_x / \#S$ for the fraction of these individuals. Individual entities carry no diversity value of their own. That is, the diversity of S is given by the diversity of the set of extant types: $\widetilde{v}(S) = v(\{x : \#S_x \neq 0\})$, where $v : 2^X \to \mathbf{R}_+$ is represented by the attribute weighting function $\lambda \geq 0$.

---

[23] For an interpretation of product diversity in terms of "potential for surprise," see Baumgärtner (2004).

Now suppose that the "effective" diversity of some set S is determined by a sampling process. Specifically, assume that from the individuals in S, a fixed number of times $k \geq 1$ some individual is randomly drawn with replacement. The replacement assumption is chosen for mathematical convenience; in some settings, a sampling without replacement may be more realistic, but we believe the difference between the two scenarios to be minor in most cases. Note that, due to the assumed replacement, the sample size may well be strictly less than k. If r_T^k denotes the probability of obtaining $T \subseteq S$ as result of sampling k times with replacement from S, then

$$v_k(S) = \sum_{T \subseteq S} r_T^k \widetilde{v}(T)$$

defines the expected diversity of the sample.[24]

It is easily seen that in fact

$$v_k(S) = \sum_{A \in 2^X} \lambda_A \left(1 - (1 - \sum_{x \in A} q_x^S)^k \right); \tag{5.1}$$

indeed, note that $1 - \sum_{x \in A} q_x^S$ is the probability that the sampled individual does not belong to A, for a single draw; since draws are independent, the probability that some individual in the sample belongs to A is $1 - (1 - \sum_{x \in A} q_x^S)^k$.

Since the expected sampled diversity $v_k(S)$ is determined by the distribution of individuals over types given by the vector $\left(q_x^S\right)_{x \in X}$, one can think of v_k in terms of an associated heterogeneity index $h = w^{k,v}$, where, for any $p \in \Delta(X)$ with rational coefficient, $w^{k,v}(p) = v_k(S)$ for any S such that $q_x^S = p_x$ for all x; (5.1) yields the following simple representation in terms of an attribute weighting expression

$$w^{k,v}(p) = \sum_{A \subseteq X} \lambda_A \left(1 - (1 - \sum_{x \in A} p_x)^k \right). \tag{5.2}$$

Note that, by Jensen's inequality, it follows immediately from (5.2) that $w^{k,v}$ is concave, hence a fortiori quasi-concave. This preference for evenness is explained naturally here by the increased chance of duplication of an individual of the same type in the sample with the prevalence of that type in the population.

Evidently, for any p, $w^{k,v}(p)$ increases with the sample size k; moreover, as the sample size becomes infinitely large, the sampled and underlying diversities becomes equal,

$$\lim_{k \to \infty} w^{k,v}(p) = v(\text{supp}\, p).$$

Thus, the sample size can be viewed as a parameter measuring the importance of rare types, thereby controlling the richness-evenness trade-off: the larger the sample, the more can one take the realization of frequent types for granted, and the more rare types matter. Since v will not in general be symmetric, neither will be $w^{k,v}$; heterogeneity will thus no longer be maximized by uniform distributions. For example, if singletons have equal value, in the hierarchical model of Figure 1, maximization of sampled diversity entails an above-average fraction of sharks (to insure against the loss of the taxon "fish" that is uniquely realized by sharks).

---

[24] The exact expression for r_T^k is of no relevance; for example, r_T^k for $\#T = k$ equals $\left(\frac{1}{\#S}\right)^k k!$.

It is instructive to consider the case of "zero similarity" that is implicitly assumed by the generalized entropy measures described above. This assumption can be made explicit here by taking the underlying diversity function to be the counting measure, $v(S) = \#S$ for all $S \subseteq X$. This yields the sampled diversity function $w^{k,\#}$ given by

$$w^{k,\#}(p) = \sum_{x \in X} \left(1 - (1 - p_x)^k\right).$$

The family of functions $\{w^{k,\#}\}$ has two points of intersection with the generalized entropy measures: the support count and the Simpson rule.[25]

This model of heterogeneity as sampled diversity invites generalizations. For example, instead of a fixed sample size, it would frequently be natural consider the sample size itself to be random. Inspired by (5.2), one can also take a more abstract route and consider indices of the form

$$h^{\phi,v}(p) = \sum_{A \subseteq X} \lambda_A \phi\left(\sum_{x \in A} p_x\right)$$

for some transformation function $\phi : [0,1] \to [0,1]$. Preference for Evenness is assured by concavity of ϕ; monotonicity of ϕ is not needed. An especially intriguing choice of ϕ is the entropic one $\phi = \phi_{ent}$, where $\phi_{ent}(q) = q \log q$. Since $h^{\phi_{ent},\#}(p)$ is the Shannon entropy of p, the indices $h^{\phi_{ent},v}$ can be viewed as *similarity-adjusted entropy indices.* Appealing as these look, their conceptual foundation is yet to be determined.

References

[1] Baumgärtner, S. (2004), Diversity as a Potential for Surprise, *mimeographed.*

[2] Baumgärtner, S. (2006), Measuring the Diversity of What? And for What Purpose? A Conceptual Comparison of Ecological and Economic Biodiversity Indices, *mimeographed.*

[3] Baumgärtner, S. (2007), Why the Measurement of Species Diversity Requires Prior Value Judgments, in: A. Kontoleon, U. Pascual and T.Swanson (eds.), *Biodiversity Economics. Principles, Methods and Applications*, Cambridge, Cambridge University Press.

[4] Berger, W.H. and F.L. Parker (1970), Diversity of Planktonic Foraminifera in Deep Sea Sediments, *Science* **168**, 1345-1347.

[5] Bervoets, S. and N. Gravel (2007), Appraising Diversity with an Ordinal Notion of Similarity: An Axiomatic Approach, *Mathematical Social Sciences* **53**, 259-273.

[6] Bossert, W., P.K. Pattanaik and Y. Xu (2003), Similarity of Options and the Measurement of Diversity, *Journal of Theoretical Politics* **15**, 405-421.

---

[25] For the former, note that $w^{\infty,\#}(p) = \#(\text{supp } p) = h_0(p)$. For the latter, note that $w^{2,\#}(p) = 2 - \sum_{x \in X} p_x^2 = 2 - \frac{1}{w_2(p)}$; $w^{2,\#}$ thus ranks distributions in the same way as h_2 does.

[7] Dixit, A. and J. Stiglitz (1977), Monopolistic competition and optimum product diversity, *American Economic Review* **67**, 297-308.

[8] Dowding, K. and M. van Hees (2007), Freedom of Choice, *this volume.*

[9] Faith, D.P. (1992), Conservation Evaluation and Phylogenetic Diversity, *Biological Conservation* **61**, 1-10.

[10] Faith, D.P. (1994), Phylogenetic Pattern and the Quantification of Organismal Biodiversity, *Phil. Trans. Roy. Soc. Lond. B* **345**, 45-58.

[11] Gärdenfors, P (1990), Induction, Conceptual Spaces and AI, *Philosophy of Science* **57**, 78-95.

[12] Gaston, K.J. (ed.) (1996), *Biodiversity. A Biology of Numbers and Difference*, Blackwell, Oxford.

[13] Gravel, N. (2007), What is diversity?, in: R. Gekker and M. van Hees (eds.), *Rational Choice and Normative Philosophy*, London, Routledge.

[14] Hill, M. (1973), Diversity and Evenness: A Unifying Notation and its Consequences, *Ecology* **54**, 427-431.

[15] Nehring, K. (1999a), Preference for Flexibility in a Savagian Framework, *Econometrica* **67**, 101-119.

[16] Nehring, K. (1999b), Diversity and the Geometry of Similarity, *mimegraphed.*

[17] Nehring, K. and C. Puppe (2002), A Theory of Diversity, *Econometrica* **70**, 1155-1198.

[18] Nehring, K. and C. Puppe (2003), Diversity and Dissimilarity in lines and hierarchies, *Mathematical Social Choice Sciences* **45**, 167-183.

[19] Nehring, K. and C. Puppe (2004a), Modelling Phylogenetic Diversity, *Resource and Energy Economics* **26**, 205-235.

[20] Nehring, K. and C. Puppe (2004b), Modelling Cost Complementarities in Terms of Joint Production, *Journal of Economic Theory* **118**, 252-264.

[21] Pattanaik, P.K. and Y. Xu (2000), On Diversity and Freedom of Choice, *Mathematical Social Choice Sciences* **40**, 123-130.

[22] Polasky, S. (ed.) (2001), *The Economics of Biodiversity Conservation*, Ashgate Publishers.

[23] Polasky, S., A. Solow and J. Broadus (1993), Searching for Uncertain Benefits and the Conservation of Biological Diversity, *Environmental and Resource Economics* **3**, 171-181.

[24] Renyi, A. (1961), On Measures of Entropy and Information, *Proceedings of the fourth Berkeley Symposium on Mathematical Statics and Probability*, Berkeley: Univercity of California Press, 547-561.

[25] ROSENBAUM, E.F. (2000), On Measuring Freedom, *Journal of Theoretical Politics* **12**, 405-421.

[26] SHORROCKS, A.F. (1984), Inequality Decomposition by Population Subgroups, *Econometrica* **52**, 1369-1386.

[27] SIMPSON, E.H. (1949), Measurement of Diversity, *Nature* **163**, 688.

[28] SOLOW, A., S. POLASKY and J. BROADUS (1993), On the Measurement of Biological Diversity, *Journal of Environmental Economics and Management* **24**, 60-68.

[29] SOLOW, A. and POLASKY, S. (1994), Measuring biological diversity, *Environmental and Ecological Statistics* **1**, 95-107.

[30] VAN DE VEL, M.L.J. (1993), *Theory of Convex Structures*, North-Holland, Amsterdam.

[31] VANE-WRIGHT, R.I., C.J. HUMPHRIES and P.H. WILLIAMS (1991), What to Protect? – Systematics and the Agony of Choice, *Biological Conservation* **55**, 235-254.

[32] VAN HEES, M. (2004), Freedom of Choice and Diversity of Options: Some Difficulties, *Social Choice and Welfare* **22**, 253-266.

[33] WEITZMAN, M. (1992), On Diversity, *Quarterly Journal of Economics* **107**, 363-405.

[34] WEITZMAN, M. (1998), The Noah's Ark Problem, *Econometrica* **66**, 1279-1298.

[35] WEITZMAN, M. (2000), Economic Profitability Versus Ecological Entropy, *Quarterly Journal of Economics* **115**, 237-263.