Département de sciences économiques

Cahier 2002-11

Similarity of Options and the Measurement of Diversity

BOSSERT, Walter
PATTANAIK, Prasanta K.
XU, Yongsheng

Université
de Montréal

Département de sciences économiques
Université de Montréal
Faculté des arts et des sciences
C.P. 6128, succursale Centre-Ville
Montréal (Québec) H3C 3J7
Canada
http://www.sceco.umontreal.ca
SCECO-information@UMontreal.CA
Téléphone : (514) 343-6539
Télécopieur : (514) 343-7221

Ce cahier a également été publié par le Centre interuniversitaire de recherche en économie quantitative (CIREQ) sous le numéro 11-2002.

This working paper was also published by the Center for Interuniversity Research in Quantitative Economics (CIREQ), under number 11-2002.

ISSN 0709-9231

CAHIER 2002-11

SIMILARITY OF OPTIONS AND THE MEASUREMENT OF DIVERSITY

Walter BOSSERT[1], Prasanta K. PATTANAIK[2] and Yongsheng XU[3]

[1] Centre de recherche et développement en économique (C.R.D.E.) / Centre interuniversitaire de recherche en économie quantitative (CIREQ) and Département de sciences économiques, Université de Montréal

[2] Department of Economics, University of California at Riverside

[3] Department of Economics, School of Policy Studies, Georgia State University

May 2002

An earlier version of this paper was presented at the International Conference on the Analysis and Measurement of Freedom in Palermo, September 2001. The authors thank Marc Fleurbaey and Bob Russell for comments and the Social Sciences and Humanities Research Council of Canada for financial support.

RÉSUMÉ

Cet article s'occupe des mesures de la diversité. Pour une notion donnée de similarité des objets dans les ensembles à mesurer, nous caractérisons des mesures de diversité descriptives.

Mots clés : diversité, similarité, mesures

ABSTRACT

This paper analyzes the measurement of the diversity of sets based on the dissimilarity of the objects contained in the set. We discuss axiomatic approaches to diversity measurement and examine the considerations underlying the application of specific measures. Our focus is on descriptive issues: rather than assuming a specific ethical position or restricting attention to properties that are appealing in specific applications, we address the foundations of the measurement issue as such in the context of diversity.

Key words : diversity, similarity, measurement

1 Introduction

Does the newly created, left-wing political party increase the diversity of political options available to the voters in a country? Will the addition of several blue-colored passenger trains to the MARTA system increase the diversity of transportation modes in the city of Atlanta? Would the extinction of giant pandas reduce the diversity of species on the planet? How does the preservation of a native American language add to the diversity of world cultures? Has the increase in concentration in the Canadian print and television media industry that occurred over the last few years reduced the diversity of opinions and viewpoints to which the general public gets exposed? The answer to each of these questions requires the *measurement of diversity.* It is thus clear that the problem of measuring diversity may arise in the context of a very broad array of issues which often figure in public discussions and debates. While diversity may be desirable in many contexts, it is not difficult to think of situations where greater diversity is not necessarily beneficial. For instance, adding a new member with very similar positions to a coalition of agents may improve the strength and cohesion of the group, which may enable it to pursue its objectives more effectively and, consequently, a reduction in diversity may be considered a good thing from the viewpoint of the group members.

In a more abstract context, in recent years, the measurement of diversity has become an increasingly important issue in the literature on the ranking of opportunity sets in terms of freedom of choice, where opportunity sets are interpreted as sets of options available to a decision maker and alternative opportunity sets are assumed to reflect possibly different amounts of freedom of choice for the decision maker.

The purpose of this paper is to present an integrated approach and a discussion of some measures of diversity that have been proposed in the literature in contributions such as Weitzman (1992), Pattanaik and Xu (2000) and Bossert, Pattanaik and Xu (2001); see also Weikard (1998) and Nehring and Puppe (2002) for discussions.

Consider the first question we posed at the beginning of this paper: does the newly created, left-wing political party increase the diversity of political options for the voters in a country? For convenience, let us call this left-wing political party l. We consider two alternative political scenarios regarding the situation before the creation of the new party. In the first scenario, to be called *situation* α, there are initially five political parties a, b, c, d and e, of which parties a to d can all reasonably be described as rightist parties with only slight differences between their platforms, and e is a centrist party. In the second scenario, to be called *situation* β, there are initially five parties f, g, h, i and j. Assume that the first four of those are leftist parties with only minor differences between their platforms, that the difference between the platform of l and that of each of f, g, h and i is also very minor and, finally, that j is a centrist party. Since l is a leftist party and is very dissimilar to the existing parties in situation α, it seems to make sense to postulate that when we add l to the set of already existing parties in this situation, the addition significantly increases the diversity of political ideologies available to the voters. On the other hand, since l is very similar to the parties that are already present in β, when we add l in this situation, such an addition does not appear to change significantly the diversity of political platforms available to the voters. For similar reasons, it also seems that the new set $\{a, b, c, d, e, l\}$ in situation α is characterized by a greater degree of diversity than the new set $\{f, g, h, i, j, l\}$ in situation β.

The gist of this simple example is that, when assessing the diversity of a set of options, the extent to which the options in the set under consideration are similar to each other should be a relevant factor. To reiterate this point, let us consider another of the questions posed earlier. In assessing the impact of preventing a native American language from becoming extinct on the diversity of cultures, it seems clear that the degree of diversity preserved depends on the language in question and its relation to other languages. If the language in question is a variation or a dialect of several other languages that continue to exist, the loss of this language would appear to entail a much less serious reduction in diversity as compared to a situation where no other surviving language is very similar to the language under consideration. Again, it is obvious that information concerning the degree of similarity between the options is of crucial importance. A moment's reflection should convince us that the answers to the other questions posed earlier also require consideration of the similarities of the options involved.

How does one view the notion of similarity in this context? One plausible view is that, for the applications we have in mind, similarity of options is a matter of *objective judgment* or *social norms.* Thus, the issue of whether two options are similar is decided by appealing to some objective judgment or social norms rather than to the opinion of individual agents. For example, in analyzing whether two political parties are similar, one can examine their differences with respect to several important issues such as tax policy, policy on education, environmental policy, national defense, etc.. If there are significant differences between the two parties with respect to the identified important issues, we can say that the two parties are not similar; if, on the other hand, the differences between the two parties with respect to the identified issues are very small, we can say that they are similar. However, we may not be satisfied with a framework that allows for only two ‘levels’ of similarity by stipulating that two alternatives are either similar or dissimilar; instead we may want to opt for an informationally richer formulation that would permit more degrees of similarity between options. In this paper, we shall consider some of these alternative frameworks.

If the notion of similarity refers to a relation between two options, the notion of diversity reflects the collective nature of similar or dissimilar options when they form a set as a whole. We view diversity as one of several criteria which may be considered revelant for the overall assessment of sets of options. As such, we do not suggest that a diversity ranking of alternative sets of options is to be identified with a measure of desirability or undesirability. Furthermore, we do not examine the ethical arguments involved in discussions as to whether diversity is desirable and, if so, to what extent. Note that, as mentioned earlier, the desirability (or lack thereof) of diversity seems to depend on the specific context under consideration, whereas the measurement of diversity as such can, as suggested above, be analyzed in a general setting. Instead, we focus on the *descriptive contents* of the notion of diversity and alternative ways of measuring it.

The remainder of the paper proceeds as follows. The next section provides, along with our basic definitions, a discussion of distance indices designed to measure the dissimilarity between individual objects in the universal set under consideration. In particular, two types of measures are introduced: ordinal distance functions and ratio-scale distance functions. Section 3 discusses the ordinal approach, and we present a characterization of a specific measure which is due to Pattanaik and Xu (2000). In section 4, we move on to ratio-scale distances and discuss the structure of a diversity measure which was introduced in Bossert, Pattanaik and Xu (2001) and which is based on ratio-scale distances. We conclude with a discussion of the relationship between this measure and a proposal due to Weitzman (1992).

2 Distance and diversity

We use X to denote the universal set of options with at least two elements. Options in X can have alternative interpretations. For example, these options may be thought of as ordinary commodity bundles. They may also be interpreted as different political parties in a country, or different species in the world, etc.. $\mathcal{K}$ is used to denote the class of all non-empty and finite subsets of X. The interpretation of the elements of $\mathcal{K}$ depends, of course, on the interpretation of the options in X.

We think of the diversity of a set as a characteristic with an *ordinal* interpretation and, thus, seek to develop ordinal measures of diversity: all we want to do is to establish a *ranking* of sets with respect to their relative diversity. Thus, we want to be able to make statements such as 'set A is more diverse than set B' but we do not attempt to perform any other comparisons, such as the comparison of differences or ratios of diversity. However, even though our interpretation of a diversity measure applied to sets of options is ordinal, this is consistent with richer informational environments when it comes to diversity comparisons of individual options. By way of analogy, note that, in social choice theory, it is typically attempted to construct a social ranking of alternatives but this ordinal interpretation of a *social* ranking certainly allows this ranking to take into account more than just ordinal and interpersonally noncomparable information at the *individual* level (for example, the *utilitarian* ordering compares any two alternatives on the basis of their respective sums of individual utilities). See Bossert and Weymark (forthcoming) for a detailed discussion of information assumptions in social choice theory.

Given this ordinal interpretation, we can employ a *binary relation* $\succeq$ on $\mathcal{K}$ as a measure of diversity. Thus, the interpretation of this relation is that we have $A \succeq B$ for any two sets A and B in $\mathcal{K}$ if and only if the diversity associated with the set A is greater than or equal to the diversity of the set B. We assume that the diversity ranking $\succeq$ is *reflexive* and *transitive.* A diversity relation on $\mathcal{K}$ is reflexive if and only if each element of $\mathcal{K}$ is at least as diverse as itself, that is, $A \succeq A$ for any set A in $\mathcal{K}$. The relation $\succeq$ is transitive if and only if chains of relative diversity are respected in the sense that if one set is at least as diverse as another and the second set is, in turn, at least as diverse as a third, then the first set is at least as diverse as the third. That is, for any three sets A, B and C in $\mathcal{K}$, $A \succeq B$ and $B \succeq C$ together imply $A \succeq C$. The diversity relations characterized in this paper are also *complete,* which means that any two distinct alternatives are comparable with respect to their diversity: for any two sets A and B with $A \neq B$, we have $A \succeq B$ or $B \succeq A$. We use the term *quasi-ordering* for a reflexive and transitive relation, and an *ordering* is a complete quasi-ordering.

Given a diversity relation $\succeq$ with the interpretation 'at least as diverse as,' we can define the associated relations 'more diverse than' and 'as diverse as.' These relations are given by the asymmetric part $\succ$ and the symmetric part $\sim$ of $\succeq$, respectively. That is, for any two sets A and B in $\mathcal{K}$, we have $A \succ B$ if and only if A is at least as diverse as B but it is not the case that B is at least as diverse as A (formally, $A \succeq B$ and not $B \succeq A$) and $A \sim B$ if and only if A is at least as diverse as B and it is also true that B is at least as diverse as A (in symbols, $A \succeq B$ and $B \succeq A$).

The purpose of this paper is to identify diversity rankings $\succeq$ with plausible properties.

In order to make reasonable progress in establishing rankings of sets in $\mathcal{K}$, it seems clear that we first need to introduce some primitive notion of similarity between the different individual options in X in terms of diversity. For example, if we have to make a judgement about whether $\{x, y\}$ is associated with more diversity than $\{z, w\}$, then, inevitably, we have to face the issue of how similar x and y are and how similar z and w are. We distinguish between two broad frameworks, depending on the informational contents of a measure of dissimilarity of options. The first one uses an *ordinal* notion of the degree of dissimilarity (or, equivalently, the distance) between any two options in X while, in the second framework, the notion of the degree of dissimilarity or distance between two options is measured by means of a *ratio scale.* We discuss both of these informational environments in turn in the following two subsections.

2.1 Ordinal distances between options

In the ordinal case, the only significant information that can be used is a ranking of distances between pairs of elements of X. We use a relation R defined on $X \times X$ for that purpose. That is, for any four alternatives x, y, z and w in X, the statement $(x, y)R(z, w)$ is interpreted as 'the distance (or dissimilarity) between x and y is greater than or equal to the distance (or dissimilarity) between z and w.' We use P for the asymmetric part of R and I for the symmetric part of R, that is, $(x, y)P(z, w)$ if and only if $(x, y)R(z, w)$ and not $(z, w)R(x, y)$, and $(x, y)I(z, w)$ if and only if $(x, y)R(z, w)$ and $(z, w)R(x, y)$. The statement $(x, y)P(z, w)$ means that the distance between x and y exceeds the distance between z and w, and $(x, y)I(z, w)$ is interpreted as the distance between x and y and the distance between z and w being equal. We assume that R is reflexive and complete. Furthermore, in line with the interpretation of this ordinal measure of distance between options, we require that: (i) $(x, y)R(z, z)$ for any three options x, y and z; and (ii) $(x, y)R(y, x)$ for any two objects x and y. The first of those two requirements says that the distance between any two options cannot be less than the distance between an option and itself. As an immediate consequence of this property, it follows that the distance between an arbitrary option and itself is the same as the distance between any other option and itself. The second of those two assumptions reflects the similarly plausible hypothesis that the distance between any two options x and y is the same as the distance between y and x.

An important special case of the ordinal framework obtains when R has exactly two equivalence classes. That is, $X \times X$ can be partitioned into two sets $\mathbf{D}$ and $\mathbf{S}$ such that

$(x, y)I(z, w)$, $(x', y')I(z', w')$ and $(x, y)P(x', y')$ for all pairs $(x, y), (z, w)$ in $\mathbf{D}$ and for all pairs $(x', y'), (z', w')$ in $\mathbf{S}$. We refer to this special case as a *simple ordinal framework* and, in that case, R is said to be a *simple ordering* on $X \times X$.

In a simple ordinal framework, a severe constraint is imposed on the relation R. If R has only two indifference classes, it is impossible to express the judgment that, for example, x and y are more dissimilar than z and w, and z and w are, in turn, more dissimilar than u and v. All that we are able to express within a simple ordinal framework is the distinction between one 'high' degree of similarity and one 'low' degree of similarity—that is, any two alternatives x and y are either similar (which is the case if the pair (x, y) is an element of $\mathbf{S}$) or dissimilar (the case where (x, y) is in $\mathbf{D}$). However, as will turn out in the following section, the simple ordinal framework allows us to obtain an interesting axiomatization of a specific diversity relation, whereas matters are still somewhat unsettled for more general ordinal measures of diversity.

2.2 Ratio-scale distances between options

If the distance between individual options is measured by means of a ratio scale, more than just ordinal information can be used in constructing a diversity ordering of sets of options. In addition to ordinal distance comparisons such as 'the distance between x and y is greater than or equal to the distance between z and w,' a ratio scale permits a much larger class of statements regarding possible relationships between distances. For instance, statements such as 'the distance between x and y is more than twice the distance between z and w' are meaningful if distance is measured by a ratio scale. This is the case because a ratio scale is *unique up to increasing linear transformations* only: the only distance measures that carry the same information as a given ratio-scale measure are positive multiples of the original measure. In contrast, if an ordinal measure of distance is employed, *all* increasing transformations of the measure carry the same information.

To give a more precise formulation of this observation, we need more than just an ordering of individual distances. In particular, we use a function d that assigns to each pair of alternatives x and y a distance $d(x, y)$. This function is called a *distance function* or a *distance index.* We assume that $d(x, x) = 0$ for all x in X, $d(x, y) > 0$ for all distinct x and y in X, and $d(x, y) = d(y, x)$ for all x and y in X. That the distance between two options is equal to zero if the two options coincide and positive if they do not is a plausible restriction. The remaining restriction is a symmetry condition, analogous to the one employed in the ordinal case: the distance between x and y is the same as the distance between y and x. If, in addition, the *triangle inequality* (which requires that $d(x, y) + d(y, z) \geq d(x, z)$ for any three options x, y and z) is satisfied, the function d is called a *metric.* Since the triangle inequality is not needed, we do not impose it in this paper; however, the results discussed here remain valid if this requirement is added.

The crucial assumption regarding d is that this function is a *ratio scale,* that is, if all values $d(x, y)$ are replaced with $\gamma d(x, y)$ for any $\gamma > 0$, the resulting measure contains the same information as d. Therefore, comparisons such as the one mentioned above are possible—if $d(x, y)$ is more than twice the distance $d(z, w)$, then $\gamma d(x, y)$ is more than twice $\gamma d(z, w)$ for any positive value of γ. That is, with an interpretation as a ratio scale, the distance index is unique up to increasing linear transformations. By contrast, if we think of a distance measure as a representation of a distance ordering R as defined in the previous subsection, the index is unique up to *arbitrary* increasing transformations, not only up to those that are linear. Ratio scales provide, therefore, a richer informational framework and allow for greater flexibility in designing diversity orderings of sets.

3 Diversity orderings based on ordinal distances

We start with the simple ordinal framework. As defined in the previous section, this means that R has two equivalence classes only and, as a consequence, any two options x and y are either similar ((x, y) is in $\mathbf{S}$) or dissimilar ((x, y) is in $\mathbf{D}$).

In order to introduce a plausible diversity ranking in this setting, we require some further definitions. A set A in $\mathcal{K}$ is *homogeneous* if and only if (x, y) is an element of $\mathbf{S}$ for all x and y in A. Therefore, a homogeneous set consists of options that are all pairwise similar to each other.

A *partition* of a set A in $\mathcal{K}$ is a collection of nonempty and disjoint subsets of A such that the union of these subsets is the set A itself. That is, a partition of a set A is a way of splitting up (partitioning) the elements of A into different subsets. A *similarity-based partition* of a set A in $\mathcal{K}$ is a partition of A such that each component of the partition is homogeneous. Because $\mathcal{K}$ is assumed to be the set of all nonempty and *finite* subsets of X, each A in $\mathcal{K}$ can be written as $A = \{a_1, \ldots, a_m\}$ where m is the finite number of elements in A. It follows that, for any A in $\mathcal{K}$, there exists at least one similarity-based partition, namely, the partition $\{\{a_1\}, \ldots, \{a_m\}\}$ where each component is a singleton.

Suppose now that the universal set of options X is finite. In this case, the number of elements contained in any nonempty subset A of X is a number between one and the number of elements in X. This implies that the number of components of any similarity-based partition of A is between one and the number of elements in X. Because there exists at least one similarity-based partition, this implies that there exists a similarity-based partition with a minimal number of components. Clearly, this similarity-based partition with a minimal number of components need not be unique. For example, if $A = \{x, y, z\}$ and $\mathbf{S}$ (the set of similar pairs) consists of the two pairs (x, y) and (y, z), there are three similarity-based partitions of A, namely, $\{\{x\}, \{y\}, \{z\}\}$, $\{\{x, y\}, \{z\}\}$ and $\{\{x\}, \{y, z\}\}$. Both $\{\{x, y\}, \{z\}\}$ and $\{\{x\}, \{y, z\}\}$ are similarity-based partitions with minimal number of components given by two.

Given the existence of a similarity-based partition with a minimal number of components for each subset A of X, we can define the following *similarity-based diversity ordering* $\succeq_s$. For all sets A and B in $\mathcal{K}$, $A \succeq_s B$ if and only if the minimal number of components of a similarity-based partition of A is greater than or equal to the minimal number of components of a similarity-based partition of B.

To illustrate this definition, consider the following example. Let $X = \{x, y, z\}$ and $\mathbf{S} = \{(x, y), (y, z)\}$. We obtain

$$\mathcal{K} = \{\{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, \{x, y, z\}\}.$$

The minimal number of components of a similarity-based partition is equal to 1 for the sets $\{x\}$, $\{y\}$, $\{z\}$, $\{x, y\}$ and $\{y, z\}$, whereas this minimal number is equal to 2 for the sets $\{x, z\}$ and $\{x, y, z\}$. Therefore, we obtain the similarity-based diversity ordering

$$\{x, z\} \sim_s \{x, y, z\} \succ_s \{x\} \sim_s \{y\} \sim_s \{z\} \sim_s \{x, y\} \sim_s \{y, z\}.$$

The reason why this ordering is of particular interest is that it can be given a plausible axiomatic justification: it is the only quasi-ordering of opportunity sets satisfying the axioms introduced below.

The first axiom is rather uncontroversial. Its counterpart in the context of ranking opportunity sets in terms of freedom of choice was introduced and discussed in Jones and Sugden (1982) and in Pattanaik and Xu (1990). It states that any two situations with no diversity at all (that is, two sets of options each of which contains a single element only) should be ranked as equally diverse by $\succeq$. The intuitive appeal of this condition in the present context is evident.

Indifference between no-diversity situations: For any two alternatives x and y in X, $\{x\} \sim \{y\}$.

Our second condition is a *monotonicity* axiom. It is concerned with diversity comparisons of an existing set A in which all the elements in A are similar to each other and an enlarged set where an option x that is outside of A is added to A. Information regarding the similarity of existing options and the option to be added is explicitly taken into account: if the added option x is similar to the (pairwise similar) options that are already present in A, it seems that the new option does not increase the amount of diversity. If, on the other hand, x is dissimilar to at least one of the alternatives in A, adding this alternative does lead to an increase in diversity. To illustrate, consider again one of the examples discussed earlier. If a newly created left-wing party l is similar to the existing parties a, b, c, d and e that are left-wing parties, then it can be argued that the set $\{a, b, c, d, e, l\}$ of parties offers the same amount of diversity of political ideologies as the set $\{a, b, c, d, e\}$; if, however, the existing parties are all rightist, then it seems plausible to argue that the set $\{a, b, c, d, e, l\}$ offers more diversity of ideologies than the set $\{a, b, c, d, e\}$. Formally, the axiom monotonicity is defined as follows.

Monotonicity: For any homogeneous set A in $\mathcal{K}$ and for any alternative x that is not in A,

(i) if (x, y) is in $\mathbf{S}$ for all y in A, then $A \cup \{x\} \sim A$;

and

(ii) if there is a y in A such that (x, y) is in $\mathbf{D}$, then $A \cup \{x\} \succ A$.

We require one more axiom to characterize the similarity-based diversity ordering. This condition deals with the response of a ranking when sets of options are merged. It is the diversity analogue of a weakening of an axiom proposed by Sen (1991) in the context of the measurement of freedom of choice. Consider two sets A and B such that A is at least as diverse as B and two sets C and D such that C is at least as diverse as D. Moreover, suppose C and D are homogeneous and A and C as well as B and D do not have any common elements. A possible composition-consistency property would require that the set obtained by merging A and C is at least as diverse as the set obtained by merging B and D. However, the axiom proposed by Pattanaik and Xu (2000) is even weaker: it requires that this property of the merged set applies only in situations where the minimal number of components of any similarity-based partition of the merged set $A \cup C$ exceeds the minimal number of components of any similarity-based partition of A. That is, the requirement only applies in situations where adding the elements of C to A actually represents a ‘real’ augmentation in terms of diversity as expressed by means of minimal similarity-based partitions. In addition, we require an analogous strict relation to be respected by set compositions of the above-described kind. This leads to the following axiom.

Composition consistency: For any two sets A and B in $\mathcal{K}$ and for any two homogeneous sets C and D in $\mathcal{K}$ such that the intersection of A and C and the intersection of B and D are empty and, furthermore, the minimal number of components of any similarity-based partition of the merged set $A \cup C$ exceeds the minimal number of components of any similarity-based partition of A,

(i) if $A \succeq B$ and $C \succeq D$, then $A \cup C \succeq B \cup D$;

and

(ii) if $A \succ B$ and $C \succeq D$, then $A \cup C \succ B \cup D$.

We now obtain the following characterization of the similarity-based diversity ordering, due to Pattanaik and Xu (2000); see their paper for a proof.

Theorem 1 *Suppose X is finite and R is a simple ordering on $X \times X$. A quasi-ordering $\succeq$ on $\mathcal{K}$ satisfies indifference between no-diversity situations, monotonicity and composition consistency if and only if $\succeq$ is equal to the similarity-based diversity ordering $\succeq_s$.*

We conclude this section with a brief discussion of the general ordinal framework, that is, we examine the construction of diversity rankings in an ordinal framework that is not necessarily a simple ordinal framework. In this case, the simple dichotomy between pairs of similar options and pairs of dissimilar options no longer applies—we may have a much richer scheme. This framework has not received much attention in the literature. We do, however, consider it worthwhile to include a few remarks on possible issues that arise in that case here.

Formally, there is an analogy between our problem of establishing a diversity ranking on $\mathcal{K}$, given the relation R defined on $X \times X$ and the problem (discussed in some parts of the existing literature on the ranking of sets) of ranking finite subsets of a set of objects, given a relation defined on this set of objects. This analogy applies if the objects in the universal set are *not* mutually exclusive; see Barberà, Bossert and Pattanaik (forthcoming) for a review of some contributions on this latter problem.

One way of bringing out this analogy is to visualize in the following fashion our problem of ranking the elements of $\mathcal{K}$ in the general ordinal framework. For all $A \in \mathcal{K}$, let Z_A denote the set of all pairs of *distinct* alternatives belonging to A. Note that Z_A is empty if A has exactly one element. The task of ranking two sets A and B by means of a ranking $\succeq$ can be interpreted as the task of ranking Z_A and Z_B. More precisely, suppose R is a relation defined on $X \times X$. One can now think of the process of deriving the relation $\succeq$ on $\mathcal{K}$ as consisting of two steps. In the first step, given the relation R, we derive a relation

$\succeq'$ defined on the class of all Z_A, where A is in $\mathcal{K}$. In the second step, the relation $\succeq$ defined on $\mathcal{K}$ is induced by $\succeq'$ in the following fashion: for any two sets A and B in $\mathcal{K}$, we let $A \succeq B$ if and only if $Z_A \succeq' Z_B$. The first step is the familiar problem (discussed in the literature on ranking sets) of ranking finite subsets of a universal set, given a relation defined on this universal set, where the options in the universal set are not mutually exclusive.

Note, however, that there is one important difference between the problem as it is usually posed in the literature on ranking sets and our problem of deriving a relation $\succeq'$ on the class of all Z_A with $A \in \mathcal{K}$ given the relation R. The existing literature typically treats the problem as one of ranking either all subsets or all non-empty subsets of a given universal set. In our case, the relation $\succeq'$ is defined merely on the class of all Z_A such that $A \in \mathcal{K}$ which, clearly, does not include all non-empty subsets of $X \times X$. Despite this difference, however, the existing results on the ranking of sets suggest some interesting questions in our context. For example, given the relation R, what are necessary and sufficient conditions for the relation $\succeq'$ to have an additive real-valued representation? That is, what are necessary and sufficient conditions under which there exists a function u such that u represents R (that is, u ranks all pairs in $X \times X$ in the same way as R) and, in addition, for any two sets A and B in $\mathcal{K}$, we have $Z_A \succeq' Z_B$ if and only if either Z_B is empty or both Z_A and Z_B are non-empty and $\sum_{(a,a') \in Z_A} u(a,a') \geq \sum_{(b,b') \in Z_B} u(b,b')$? This question is of interest because, as we have noted earlier, the relation $\succeq'$ can be used to induce, in a straightforward fashion, the relation $\succeq$ that we are ultimately interested in. Given the formal difference noted above, the answer to this question does not follow immediately from the corresponding results in the existing literature. This is an example of a problem that merits further investigation.

4 Diversity orderings based on ratio-scale distances

We now consider a framework that allows us to use more than just ordinal information concerning the dissimilarity between individual objects in X. In particular, we use a distance function d as introduced in Section 2 that we interpret as a ratio scale—that is, the function is unique up to increasing linear transformations. For any two objects x and y in X, the value $d(x,y)$ is the distance between x and y. Given this measure of distance between objects, we can define the notion of distance between an object x and a set A as

$$d(x,A) = \begin{cases} 0 & \text{if } A = \{x\}, \\ \min\{d(x,y) \mid y \in A \setminus \{x\}\} & \text{if } A \neq \{x\}. \end{cases}$$

According to this definition, the distance betweeen a singleton set and its constituent element is zero, and the distance between any other combination of a set and an element is positive. Our results would not change if we amended this definition to one that is more commonly used and assigns a zero distance whenever the object under consideration is an element of the corresponding set; we have chosen the above formulation merely because it is more convenient for the exposition of our analysis.

The idea underlying the diversity measure we propose in this section is to aggregate the distances between objects and other elements (if any) in a set A in a systematic and plausible way. To do so, (at least) two points need to be observed: we have to avoid the 'multiple-counting' of distances and, on the other hand, we want to ensure that the distances between very dissimilar objects are accounted for properly. To take due consideration of those requirements, we employ an iterative procedure for each set, which ensures that multiple-counting is ruled out. We keep track of the distance between a specific element of a set and the other elements of the set (if there are any), then eliminate this element, and repeat the procedure until the set is exhausted. This procedure is well-defined because we only compare finite sets. What is important for the specific properties of this procedure is, of course, the choice of the element to be eliminated at each stage of the iteration. To ensure that substantial dissimilarities are accounted for in a suitable manner, we employ a *leximin* elimination criterion with respect to the minimal distance between an object in a set and the remaining elements of the set.

In more precise terms, the iterative procedure used to define a measure of diversity can be described as follows. Consider any set A in $\mathcal{K}$. For any x in A, we record the distances between x and all elements of A (including x itself) in a vector $\delta_A(x)$. To illustrate, consider the following example. Suppose we have a set $A = \{x, y, z, w\}$ and the distances between the objects in A are given by $d(x, y) = 2$, $d(x, z) = 1$, $d(x, w) = 2$, $d(y, z) = 3$, $d(y, w) = 2$ and $d(z, w) = 4$. Note that these definitions completely specify all pairwise distances between objects in A; all distances that are not explicitly written out are obtained by the symmetry of the function d and the property that the distance between any object and itself is equal to zero. The vector of distances between x and all objects in A (including x itself) is given by $\delta_A(x) = (0, 2, 1, 2)$. Analogously, we obtain $\delta_A(y) = (2, 0, 3, 2)$, $\delta_A(z) = (1, 3, 0, 4)$ and $\delta_A(w) = (2, 2, 4, 0)$.

After having obtained these vectors for each object in A, we compare them according to the leximin criterion. That is, we begin with the smallest component of each of those vectors and compare their smallest distances. If there is a unique option that has the smallest minimal distance among all minimal distances, we use it as the first object in our elimination procedure for the set A and call it a_1. If there are several objects in A with a smallest minimal component of their respective vector as constructed above, we compare their next-to-minimal components, and so on. If this process yields a unique object after a finite number of steps, we call this object a_1. If we end up with more than one object after this lexicographic procedure, it does not matter which one we pick and we select an arbitrary one among them and call it a_1. Because each set A has a finite number of elements, this procedure is well-defined and terminates after a finite number of comparisons.

We now record the value of $d(a_1, A)$, eliminate a_1 and repeat the procedure with A replaced by $A \setminus \{a_1\}$. That is, we find an element a_2 in $A \setminus \{a_1\}$ whose vector of distances is a smallest vector according to the leximin criterion, record the value of $d(a_2, A \setminus \{a_1\})$, and eliminate a_2 from $A \setminus \{a_1\}$ in order to repeat the procedure with $A \setminus \{a_1\}$ replaced by $A \setminus \{a_1, a_2\}$. Again appealing to the finiteness of A, we conclude that this procedure terminates after a finite number of iterations, and we have obtained the distances $d(a_1, A), d(a_2, A \setminus \{a_1\}), \ldots, d(a_{m-1}, A \setminus \{a_1, \ldots, a_{m-2}\}), d(a_m, \{a_m\})$, where m is the number of elements in A.

We use the above example again to illustrate the iterative procedure. According to the leximin criterion, x is associated with the unique smallest vector of distances. Its minimal component is equal to 0, as is the minimal component of each of the other objects. The next-to-smallest distance in $\delta_A(x)$ is equal to 1, and so is the next-to-smallest distance in $\delta_A(z)$. On the other hand, the corresponding value for $\delta_A(y)$ and for $\delta_A(w)$ is 2. Thus, we move on to compare the third-smallest components of $\delta_A(x)$ and $\delta_A(z)$. We obtain values of 2 for x and 3 for z and, thus, our object a_1 is given by x. The value of $d(a_1, A)$ is therefore equal to $d(x, A) = 1$. Now we repeat the procedure with $A \setminus \{a_1\} = A \setminus \{x\} = \{y, z, w\}$ instead of A. We obtain $\delta_{A \setminus \{x\}}(y) = (0, 3, 2)$, $\delta_{A \setminus \{x\}}(z) = (3, 0, 4)$ and $\delta_{A \setminus \{x\}}(w) = (2, 4, 0)$. The smallest of these vectors according to the leximin criterion is that corresponding to y and, thus, we set $a_2 = y$ and obtain $d(a_2, A \setminus \{a_1\}) = d(y, A \setminus \{x\}) = 2$. This leaves us with the set $A \setminus \{x, y\} = \{z, w\}$ and the vectors of distances $\delta_{A \setminus \{x,y\}}(z) = (0, 4)$ and $\delta_{A \setminus \{x,y\}}(w) = (4, 0)$. Clearly, either z or w can be chosen for a_3 now, and we obtain $d(a_3, A \setminus \{a_1, a_2\}) = d(z, A \setminus \{x, y\})$ or $d(a_3, A \setminus \{a_1, a_2\}) = d(w, A \setminus \{x, y\})$ which, in either case, yields $d(a_3, A \setminus \{a_1, a_2\}) = 4$. Obviously, $d(a_4, A \setminus \{a_1, a_2, a_3\}) = d(a_4, \{a_4\}) = 0$ for either choice of a_3 (and, thus, of a_4).

Finally, we define the *leximin diversity ordering* $\succeq_\ell$ as follows. For any two sets A and B in $\mathcal{K}$, we have $A \succeq_\ell B$ if and only if the sum $d(a_1, A) + \ldots + d(a_m, \{a_m\})$ is greater than or equal to the sum $d(b_1, B) + \ldots + d(b_n, \{b_n\})$, where n is the number of elements in B

and the $b_1, \ldots, b_n$ are obtained for B in the same way the $a_1, \ldots, a_m$ are obtained for A.

Going back to our example, the sum $d(a_1, A) + d(a_2, A \setminus \{a_1\}) + d(a_3, A \setminus \{a_1, a_2\}) + d(a_4, A \setminus \{a_1, a_2, a_3\})$ is given by

$$d(x, \{x, y, z, w\}) + d(y, \{y, z, w\}) + d(w, \{z, w\}) + d(z, \{z\}) = 1 + 2 + 4 = 7.$$

The ranking of A and any other set B is now determined by computing the corresponding sum for B and comparing it to the value of 7 obtained for A.

We now turn to a characterization of $\succeq_\ell$. The first axiom we use is another monotonicity condition. It applies to diversity comparisons of sets with at most two elements, which is why we refer to it as *simple* monotonicity. The axiom requires that the ranking of two sets with at most two elements each is determined by the individual distance between the two elements. This is a very plausible requirement: if there is a single individual distance only within each of two sets, then this single distance should be considered the aggregate distance for each set as well in comparing the two.

Simple monotonicity: For all objects x, y, z, w in X, $\{x, y\} \succeq \{z, w\}$ if and only if $d(x, y) \geq d(z, w)$.

The next axiom expresses an invariance property with respect to certain additions to sets. Consider two sets A and B, an object x that is not in A and an object y that is not in B. The axiom requires that, under some circumstances, the relative ranking of A and B according to $\succeq$ is unchanged if x is added to A and y is added to B. To specify the conditions under which the axiom applies, suppose that, according to the leximin criterion, the vector of distances associated with x is a smallest element within the set that consists of A augmented by x and, analogously, the distance vector of y within $B \cup \{y\}$ is a smallest element. Finally, suppose that the distance between x and A is the same as the distance between y and B. If all those requirements are satisfied, the axiom *independence* requires the relative ranking of A and B to be the same as the relative ranking of $A \cup \{x\}$ and $B \cup \{y\}$. The underlying idea is that simultaneous additions of elements to two sets such as those described above do not change the relative diversity of the two sets.

Independence: For any two sets A and B in $\mathcal{K}$, for any two objects x and y such that x is not in A and y is not in B, if the vector of distances for x within $A \cup \{x\}$ is minimal, the vector of distances for y within $B \cup \{y\}$ is minimal and $d(x, A) = d(y, B)$, then $A \succeq B$ if and only if $A \cup \{x\} \succeq B \cup \{y\}$.

For the formulation of our last axiom, we introduce the notion of a *link option.* Consider two two-element sets $\{x, y\}$ and $\{z, w\}$ such that x and y are distinct, and so are z and w. An object u that is different from z and from w is a *link option of* $\{z, w\}$ *relative to* $\{x, y\}$ if it has the following properties. The object u has to be 'between' z and w in the sense that neither the distance between u and z nor the distance between u and w exceeds the distance between z and w. Furthermore, the distance between x and y must be equal to the sum of the distances between z and w and between u and $\{z, w\}$. That is, u provides a 'link' between z and w that reproduces the distance between x and y. This notion of a link option is related to Weitzman's (1992) *link property.* Weitzman postulates, for every set A in $\mathcal{K}$, the existence of a 'link species' defined as an option x of A such that the value of a representation of the diversity ordering at A is equal to the sum of the value of this representation at $A \setminus \{x\}$ and the distance between x and $A \setminus \{x\}$. Weitzman's requirement is not suitable for our context, however: its formulation requires *more* than just an ordinal interpretation of a diversity measure and, thus, cannot be expressed in terms of a diversity ordering $\succeq$. For that reason, we believe that our notion of a link element is easier to justify.

The axiom *link indifference* requires that adding a link option of a set $\{z, w\}$ relative to a set $\{x, y\}$ leads to a set that is indifferent to $\{x, y\}$, provided that the elements in each set are distinct and the set $\{x, y\}$ is more diverse than the set $\{z, w\}$. This requirement states that the addition of an object to a set for which it is a link option offsets the higher diversity (provided that this diversity actually *is* higher) of a set with respect to which the object is a link option. This requirement is in line with the definition and interpretation of a link option.

Link indifference: For all objects x, y, z, w, u in X such that $x \neq y$ and z, w and u are pairwise distinct, if $\{x, y\} \succ \{z, w\}$ and u is a link option of $\{z, w\}$ relative to $\{x, y\}$, then $\{x, y\} \sim \{z, w, u\}$.

These axioms can be used to provide a characterization of the leximin diversity ordering. We need a regularity requirement that is, in essence, a richness property regarding the universal set X and the distance function d. The presence of such a condition is required because without it, the link-indifference property would not have any bite. We obtain the following result which is due to Bossert, Pattanaik and Xu (2001); see that paper for a proof.

Theorem 2 *Suppose X is an infinite universal set. Furthermore, suppose X and d are such that, for all numbers s and t with $t \leq s$, there exist options x, y and z such that*

$t = d(x,y) \leq d(x,z) \leq d(y,z) = s$. A diversity ordering $\succeq$ satisfies simple monotonicity, independence and link indifference if and only if $\succeq = \succeq_\ell$.

It is worth noting the connection between our diversity measure $\succeq_\ell$ and the one introduced by Weitzman (1992). Weitzman proposes ameasure that is defined implicitly as the solution of a recursive programming problem; see Weitzman (1992) and Bossert, Pattanaik and Xu (2001) for a precise definition. We think that, owing to this implicit way of defining the measure, the calculation of the measure as well as its properties are not very transparent. Thus, it is a useful observation that his measure coincides with ours, the calculation of which can be done explicitly by means of the iterative procedure discussed earlier. As a consequence, the above characterization result can also be seen as a further justification of Weitzman's approach.

References

Barberà, S., W. Bossert, and P.K. Pattanaik, "Ranking sets of objects", in: Barberà, S., P.J. Hammond and C. Seidl (eds.), *Handbook of Utility Theory, Vol. II: Extensions,* Kluwer Academic Publishers, Dordrecht, forthcoming.

Bossert, W., P.K. Pattanaik and Y. Xu, "The measurement of diversity", *Discussion Paper* 01-21, Department of Economics, University of California at Riverside, 2001.

Bossert, W. and J.A. Weymark, "Utility in social choice", in: Barberà, S., P.J. Hammond and C. Seidl (eds.), *Handbook of Utility Theory, Vol. II: Extensions,* Kluwer Academic Publishers, Dordrecht, forthcoming.

Jones, P. and R. Sugden, "Evaluating choice", *International Review of Law and Economics,* 2 (1982), 47–65.

Nehring, K. and C. Puppe, "A theory of diversity", *Econometrica,* 70 (2002), 1155-1198.

Pattanaik, P.K. and Y. Xu, "On ranking opportunity sets in terms of freedom of choice", *Recherches Economiques de Louvain,* 56 (1990), 383–390.

Pattanaik, P.K. and Y. Xu, "On diversity and freedom of choice", *Mathematical Social Sciences,* 40 (2000), 123–130.

Sen, A. K., "Welfare, preference and freedom", *Journal of Econometrics,* 50 (1991), 15–29.

Weikard, H.-P., “On the measurement of diversity”, *Discussion Paper* 9801, Institute of Public Economics, University of Graz, 1998.

Weitzman, M., “On diversity”, *Quarterly Journal of Economics,* 107 (1992), 363–405.