



HAL
open science

Relational diversity

Hans-Rolf Gregorius

► **To cite this version:**

Hans-Rolf Gregorius. Relational diversity. Journal of Theoretical Biology, 2009, 257 (1), pp.150.
10.1016/j.jtbi.2008.11.009 . hal-00554533

HAL Id: hal-00554533

<https://hal.science/hal-00554533>

Submitted on 11 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Relational diversity

Hans-Rolf Gregorius

PII: S0022-5193(08)00585-7
DOI: doi:10.1016/j.jtbi.2008.11.009
Reference: YJTBI5361

To appear in: *Journal of Theoretical Biology*

Received date: 18 August 2008
Revised date: 23 October 2008
Accepted date: 4 November 2008

Cite this article as: Hans-Rolf Gregorius, Relational diversity, *Journal of Theoretical Biology* (2008), doi:[10.1016/j.jtbi.2008.11.009](https://doi.org/10.1016/j.jtbi.2008.11.009)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jtbi

Relational diversity

HANS-ROLF GREGORIUS^{1,2}

¹Institut für Populations- und ökologische Genetik
Am Pfungstanger 58, 37075 Göttingen, Germany

²Abteilung Forstgenetik und Forstpflanzenzüchtung
Büsgenweg 2, 37077 Göttingen, Germany

Abstract

In biology, the measurement of diversity traditionally focusses on reporting numbers of unambiguously distinguishable types, thus referring to qualitative (discontinuously varying) traits. Inclusion of frequencies or other weights has produced a large variety of diversity indices. Quantitative (continuously varying) traits do not readily fit into this perspective. In fact, in the context of quantitative traits, the concept of diversity is not always clearly distinguished from the (statistical) notion of dispersion. In many cases the ambiguity even extends to qualitative traits. This is at variance with the broad spectrum of diversity issues ranging e.g. from ecological and genetic aspects of diversity to functional, structural, systematic or evolutionary (including phylogenetic) aspects. In view of the urgent need for a more consistent perspective, it is called to attention that all of these aspects, whether of qualitative or quantitative nature, can be gathered under the common roof of binary relations (for qualitative traits two objects are related, for example, if they share the same trait state). A comprehensive concept of (relational) diversity can be developed in two steps: (1) determine the number of unrelated pairs of objects among all admissible pairs as a measure of implicit (relative) diversity, (2) invoke the concept of effective number to transform the implicit measure of diversity into an explicit (absolute) measure. The transformation operates by equating the observed implicit diversity to the implicit diversity obtained for the ideal model of an equivalence relation with classes of equal size. The number of these classes specifies the effective number as an explicit measure of diversity. The wealth of problems that can be treated from this unified perspective is briefly addressed by classifying and interpreting established diversity indices in the light of relational diversity. Desirable applications to the above-mentioned aspects are specified with the help of types of relations such as order, hierarchical, and tree relations. Corresponding biological issues including taxonomic community diversity, mating system, food web, sociological, cladistic and phylogenetic, or hypercycle diversity are suggested for future consideration.

Introduction

Especially when considering adaptational processes, it becomes obvious that diversity is the prerequisite for the functioning of these processes and that maintenance of the processes requires maintenance of diversity. Diversity thus plays a fundamental role in the closed cause-effect cycles that serve the sustainment of biological systems. The quality and the quantity of diversity that is to be preserved may however vary considerably

among systems and their functions. The latter explains why most of the debate about appropriate methods of measuring biological diversity is concerned with aspects of the quality and quantity of diversity. Aspects of quality may be of ecological, functional, structural, systematic or evolutionary kind as well as of a practical kind relevant in the regulation and conservation of natural and artificial biological systems. Aspects of quantity are chiefly addressed via diversity indices designed to quantify in various ways the aforementioned aspects of quality. An impression of more recently addressed problems in this context can be gained from a series of articles published 2006 in *Acta Biotheoretica* Volume 54 Number 2 (Myers & Patil, Podani, Sarkar, Ricotta).

For most applications in biology, the measurement of diversity refers to the distribution of qualitative traits displayed by the members of a specified collection. Diversity measurement is then primarily focussed on finding equivalents of the number of different trait states or types (frequently called “effective number”). According to Jost (2006), the “effective number ... is the key to a unified and intuitive interpretation of diversity”. There seems to be general agreement that this elementary focus on numbers of types identifies the basic criterion to be fulfilled by diversity measures (it also reflects appropriately the definition of diversity as variety or multiformity given in most dictionaries). Depending on the field of interest, types may be species or other systematic, sociological or ecological categories, alleles, single- or multi-locus genotypes, etc.

Counting types in a straightforward manner of course requires that these types are unambiguously distinguishable, and this usually applies only to qualitative traits. With reference to the objects that display the types, this requirement implies distinction of objects in a strictly binary (or qualitative) fashion, i.e. by stating the presence or the absence of difference in the trait states of two objects without any consideration of variable degrees or amounts of difference. In a basic sense, objects are therefore considered to be related by sharing their trait state, and they are otherwise considered to be unrelated. Under this premise, the number of trait states is determined by the unrelated objects. Relationships other than those characterized by identity and non-identity are not part of this perspective.

Quantitative traits, however, are also sometimes analyzed in terms of their diversity (see e.g. Rao, 1982, for one of the earliest and most comprehensive treatments, also see Shimatani, 2001), even though the underlying concept of diversity is not always clearly distinguished from the concept of dispersion which determines the measures of variation commonly used in statistics. Even though some measures of dispersion can be transformed into effective numbers (as is implicit in the paper of Rao, 1982, and as was highlighted by Jost, 2006), this is not part of their conceptual underpinning (for a discussion of other approaches based on quantitative traits such as that of Weitzman, 1992, see Gregorius & Gillet, 2008).

In any case, whether qualitative or quantitative traits are considered, the basic idea of diversity is the notion of difference e.g. in appearance or function between pairs of objects. The terms difference, similarity or identity are used in this context to express relations between pairs of objects that might be defined by degrees to which the objects

are different, similar or identical. Yet, continuously varying differences lend themselves neither to the separation of distinct types nor to the determination of effective numbers, as is inherent in the concept of diversity.

It is nevertheless possible to include quantitative traits into the measurement of diversity. One way is suggested by choosing a threshold difference as a basis for binary distinction. Exceeding the threshold is then the criterion for stating distinguishability. Accordingly, objects are considered to be related if they differ by not more than the threshold. From this perspective, relatedness can be addressed at various levels of resolution by varying the threshold difference. Each level of resolution can then be assigned its appropriate measure of diversity (see Gregorius & Gillet, 2008).

The method of using thresholds to define relations can also be applied to weighted graphs. Examples range from evolutionary ecology to gene expression systems (see e.g. Wootton, 2005, or Xu et al., 2002). In such graphs minimum path lengths are defined from one vertex (object) to another. Given this information, one vertex can be considered to be related to another if the minimum path length does not exceed a specified threshold. Note that this is a directed relation, since paths may exist in only one direction. If paths exist in both directions they may differ in length, so that the threshold condition for relatedness may be obeyed in one but not in the other direction.

Other familiar examples of relations come from forms of biological interaction in which individuals or species communicate through positive (symbiotic) or negative (antagonistic) interactions or are linked through food chains or sociological associations. In contrast to the examples on quantitative traits, these kinds of relations do not appear to require any measurement of difference between objects. This, however, may be misleading, since there is always the basic prerequisite that the presence of the respective form of interaction can at least be unambiguously distinguished from its absence.

Structural diversity is another topic that has received considerable recent interest. In fact, the term “structure” is most closely associated with the notion of relation, since it explicitly aims at the characterization of interrelationships between parts in an entity (in which most dictionaries seem to agree). A distinction has to be made however between the diversity of structures in the sense of differences between structures (and thus relations), and the diversity observable within a particular type of structure. It is the latter that will be addressed in this paper, since the former is rather a problem of differentiation between than diversity within structures. Probably in the majority of studies, the term structural diversity is used in connection with quantitative traits, and by this have to deal with the conceptual difference between the measurement of dispersion and diversity (for a typical example see Staudhammer & LeMay, 2001).

Extrapolation of the above examples of relations directly points to the possibility that general relations between members of a collection can be considered as a source of diversity. By this, a wide spectrum of current methods can be traced back to a common principle, and new subjects can be suggested. In fact, relations will be shown to have the potential for a unified approach to the measurement of diversity, just as the average distance considered by Rao (1982) has for the measurement of dispersion. Different

qualities of diversity can be contributed to different types of relation, and measures of the quantity of diversity (called relational diversity) will be seen to follow naturally from counting relations. Herewith the principle of interpreting diversity in terms of numbers of trait states or types will be ensured with the help of a generalized concept of effective number.

Formal representation

Let us start from the general mathematical definition of a binary relation as a subset R from the cross-product $K \times K$, where K is any collection of objects. Thus an object $x \in K$ is related to an object $y \in K$ if the pair (x, y) belongs to R , i.e. $(x, y) \in R$. As was indicated above, there are many situations of basic biological interest, where relations between objects are not solely aimed at the measurement of difference between objects. It is therefore justified to treat the term relatedness in a very comprehensive manner. However, since the focus is on diversity, relatedness of an object to itself (reflexivity) is not an issue. This suggests replacement of $K \times K$ by the “off-diagonal” cross-product $C_K := \{(x, y) \mid x \in K, y \in K, x \neq y\}$ and R by the “off-diagonal” relation $R^* := R \cap C_K$ in considerations of relational diversity (the wording “off-diagonal” refers to the typical illustration of binary relations in the form of quadratic matrices; see Figure 1 for a demonstration of the matrix representation and the corresponding graph representation of relations).

Apparently, if each member of a collection is related to each other of its members, this collection would be considered as displaying no structure because all objects are meshed. Displaying no structure can in turn be taken as a synonym for homogeneity and thus for the absence of diversity. Thus, the heterogeneity or diversity of a collection would increase with the number of unrelated objects. A probably more familiar illustration can be arrived at by considering a subdivision of a collection into clusters such that members of the same cluster are more frequently related among each other than to members of other clusters. Diversity increases as the differences between clusters become more expressed by fewer relations between clusters.

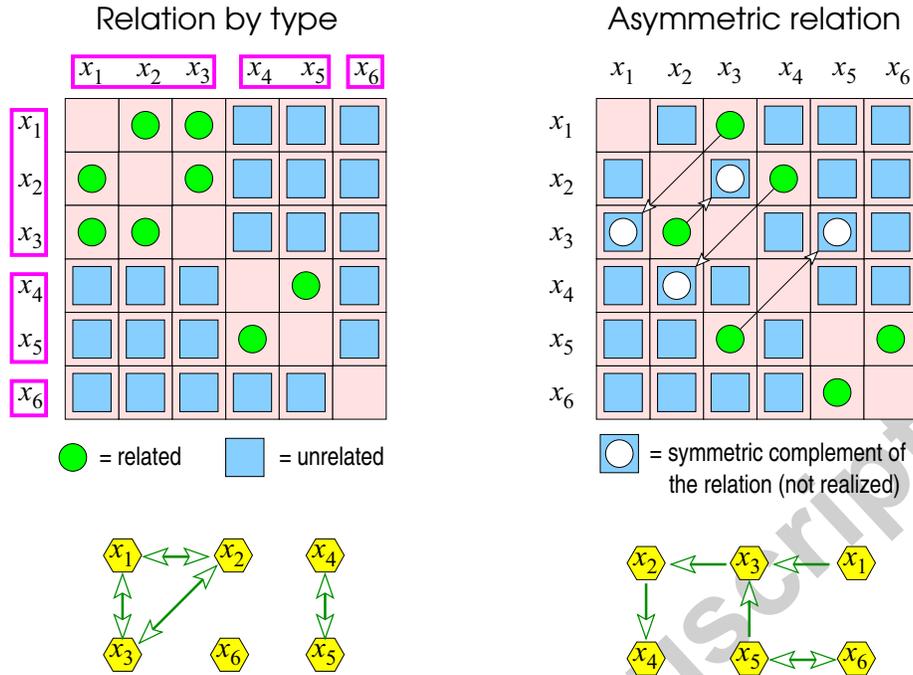
In other words, relational diversity should correspond to the number of pairs $(x, y) \notin R$ among all pairs $(x, y) \in C_K$. A straightforward *relative measure of relational diversity* would therefore be given by

$$\mathcal{D} := \frac{|C_K| - |R^*|}{|C_K|} \quad (1a)$$

where $|R^*|$ and $|C_K|$ is the number of pairs in R^* and C_K , respectively. For a number n of objects in K one obtains $|C_K| = n \cdot (n - 1)$ (see Figure 1 for an example). An alternative characterization of the measure \mathcal{D} , which accounts for collections of infinite size, is provided by *the probability of sampling without replacement two objects of which the first is not related to the second*.

In the complete absence of relations among the objects, $|R^*| = 0$ and thus $\mathcal{D} = 1$. Each object is relationally isolated in this situation as is typical of a totally divers collection. At the other extreme, where each member is related to each other member of the

Figure 1



Off-diagonal matrix and graph representation of a binary relation between the $n = 6$ objects forming the collection $K = \{x_1, \dots, x_6\}$.

Left panel: Relations are symmetric and exist between all objects of the same type but not between objects of different type. The collection is divided into three subsets corresponding to three types, $\{x_1, x_2, x_3\}$, $\{x_4, x_5\}$, and $\{x_6\}$. The off-diagonal relation R^* consists of 8 relations; relative relational diversity $\mathcal{D} = 0.73$; explicit relational diversity (effective number) $v = 2.57$.

Right panel: Relations are asymmetric with one exception. Off-diagonal relation $R^* = \{(x_1, x_3), (x_2, x_4), (x_3, x_2), (x_5, x_3), (x_5, x_6), (x_6, x_5)\}$; relative relational diversity $\mathcal{D} = 0.80$; explicit relational diversity (effective number) $v = 3.00$.

collection, $|R^*| = |C_K|$ and consequently $\mathcal{D} = 0$, indicating complete homogeneity and thus the absence of diversity. These characteristics of \mathcal{D} are in accordance with the above remarks on the diversity of collections. It also becomes apparent that relationally isolated objects cannot just be excluded from considerations of the diversity of a collection.

So far, relatedness was intuitively referred to similarity, and the notion of heterogeneity was accordingly associated with difference. However, there may be cases, where, for example, special preferences for mating partners or other types of biological cooperation are determined by dissimilarity rather than similarity. In such cases relationships between objects are primarily described by sufficient difference (e.g. avoidance of inbreeding or symbiosis via avoidance of competition). Homogeneity would thus be characterized by the fact that all individuals would equally avoid cooperation with like types. Even though this seems to contradict the above concept of diversity, it yet is in complete accordance with the conceptual basis, since it simply reverses the argument.

In fact making use of the complement R^c of a relation R , equation (1a) still applies. The *complementary relation* is here defined as usual by

$$R^c := \{(x, y) \mid x \in K, y \in K, (x, y) \notin R\},$$

so that in equation (1a) R is to be replaced by R^c to yield the pertinent measure of diversity (the *-notation retains its meaning and C_K remains the same). Under this premise it follows consistently that the relative relational diversity of the complementary relation equals $1 - \mathcal{D}$. This merely emphasizes the obvious fact that *homogeneity and heterogeneity are complementary and exchangeable notions in terms of relations*, and that therefore they must be *a priori* specified in order to avoid conceptual confusion.

Relations between objects and between types of objects

In many studies, relations are not primarily defined between individual objects but rather between the types represented by the objects. In a trioecious sexual system, for example, there are three sexual types (female, male and cosexual referred to as f, m, c) between which four symmetrical mating compatibility relations exist: (f, m) , (f, c) , (m, c) , (c, c) . If the sexual types would be considered as objects, the “diagonal” relation (c, c) would have to be discarded for an assessment of relational diversity. By equation (1a) one arrives at $\mathcal{D} = 0$ (since $n = 3$ and $C_K = |R^*| = 6$) and thus the absence of relative relational diversity. This makes sense, since each of the three types is related with each other type, which would constitute a situation of relational homogeneity.

At the other extreme, if diversity is solely conceived in terms of the number of types, then no object (type) is related to any other object (type), with the result that $\mathcal{D} = 1$. This situation is commonly addressed by the term “richness” such as in “species richness” or “allelic richness”. It will later on be returned to in connection with effective numbers.

In real collections, however, types occur through the objects representing them, so that relations between types carry over to relations between objects. Because no further information about relationships between objects other than between their types exists, one object is considered to be related to another if and only if the type represented by the first object is related to the type represented by the other object. This transfer of relations between trait states to relations between objects will be referred to as the *TO-perspective* in the following.

The TO-perspective introduces frequencies into the notion of relations as becomes apparent from the fact that the i -th type may be represented by several objects, n_i say, so that $\sum_i n_i = n$. Provided the i -th type is related to the j -th and $i \neq j$, there are then $n_i \cdot n_j$ relations between the corresponding objects. For $i = j$, there are $n_i \cdot (n_i - 1)$ relations between objects of the i -th type (provided this type is related to itself). Note that within such a group all objects are mutually related.

We are now concerned with two relations, one, R_T , refers to a set of types K_T , and the other, R , results from this by the above-described transfer to relations between objects from K . The transfer implies that each pair $(i, j) \in R_T$ can be assigned a frequency of $n_i \cdot n_j$ for $i \neq j$ and of $n_i \cdot (n_i - 1)$ for $i = j$. The relative measure \mathcal{D} of relational diversity,

which is defined for the relations between objects, can now be specified by considering that $|R^*|$ in equation (1a) takes the form $|R^*| = \sum_{(i,j) \in R_T} n_i \cdot (n_j - \delta_{i,j})$, where $\delta_{i,j}$ denotes the Kronecker symbol (i.e. $\delta_{i,j} = 0$ for $i \neq j$ and $\delta_{i,i} = 1$). In the special case, where all types are related to themselves but to no other, one obtains $|R^*| = \sum_i n_i \cdot (n_i - 1)$ so that $\mathcal{D} = 1 - \sum_i n_i \cdot (n_i - 1) / (n \cdot (n - 1))$, which is well-known as Simpson's index of diversity.

In other studies, relations are observed among objects without any prior reference to traits. Examples are provided by studies of mating systems where the fusion of gametes or copulation constitutes the event that defines relations among potential parents irrespective of any trait characteristics of the individuals. Encounters of individuals leading to positive or negative interactions provide another wide variety of relations that are observed solely on the level of individuals. Apparently, such observations do not allow for any suggestion as to the potential causes that brought about the relation. In fact, cause-effect analyses cannot be performed unless the objects are characterized by traits for which appropriate relations can be hypothesized and tested.

As opposed to the TO-perspective, relations between objects are now verified as to whether they are reflected by the states of a trait under consideration. In other words, relations between objects are transferred to relations between types. This change in perspective will be referred to as the *OT-perspective* in the following. The OT-perspective has to take account of constraints that are due to the nature of the objects. For example, it may be impossible from the outset to find types for which either each object of one type is related to each object of the other type, or for which no object of one type is related to an object of the other type, as was part of the TO-perspective. This is relevant whenever objects cannot realize multiple relations as is the case in mating system studies in which gametes are the objects between which relations are defined by their fusion (each gamete can fuse only once).

To arrive at a comprehensive representation of the OT-perspective, resume the idea of a trait as a function T that assigns to each object x a state $T(x)$. One thus has a mapping $T: K \rightarrow K_T$ with $K_T = \{T(x) \mid x \in K\}$. The two relations to be considered are again R for the objects and R_T for the trait states. Since R_T is conceived of as a transfer of R via T , the two relations correspond to each other in that

- (a) if $(x, y) \in R$ then $(T(x), T(y)) \in R_T$, and
- (b) if $(i, j) \in R_T$ then there exists at least one $x \in K$ with $T(x) = i$ and at least one $y \in K$ with $T(y) = j$ so that $(x, y) \in R$.

Note that $(x, y) \notin R$ does not rule out the possibility that $(T(x), T(y)) \in R_T$, since objects other than x and y may be related and show the same types.

Conditions (a) and (b) determine an obvious rule of transferring any relation R via a trait T to a relation R_T between the trait states. The transfer is defined by: $i \in K_T$ is related to $j \in K_T$ (i.e. $(i, j) \in R_T$) if there exist $x \in K$ and $y \in K$ such that $T(x) = i$, $T(y) = j$ and $(x, y) \in R$. By this rule, any trait $T: K \rightarrow K_T$ carries the relation R on K in a unique manner to a relation R_T on K_T . Special features of the relation R can in this way be studied with the help of appropriate traits.

For example, consider the above-mentioned mating system study in which relations between gametes are defined by their fusion, and where gametes are characterized by suspected incompatibility types of their producers. Thus the objects of interest are gametes, which are (symmetrically) related by fusion events and which are assigned the incompatibility types of their producers as trait states. Two incompatibility types are then related if at least one pair of gametes that represent the two types is found to have fused. If in particular the incompatibility types represent a system of self-incompatibility, then one expects that self-relatedness (reflexivity) is excluded from K_T . Hence, if there exists $(x, y) \in R$ with $T(x) = T(y)$, this would contradict the assumption that the observed mating relations follow a system of self-incompatibility. The example demonstrates the use of traits for testing model hypotheses on observed relations under the OT-perspective. This does of course not apply to the TO-perspective, since there the primary observation relies on relations between types rather than objects.

As in the TO-perspective, the relations between types can again be assigned the frequencies of their corresponding relations from R . Thus with each $(i, j) \in R_T$ there is associated the number $n_{i,j}$ of relations $(x, y) \in R$ with $T(x) = i$ and $T(y) = j$. This number can never exceed the number $n_i \cdot (n_j - \delta_{i,j})$ that results under the TO-perspective. Consequently, in the relative relational diversity as specified in equation (1a) we now have $|R^*| = \sum_{(i,j) \in R_T} n_{i,j}$, so that \mathcal{D} under the OT-perspective, can never be greater than under the TO-perspective.

Relations associated with weights

In the last section it was shown that for relations based on the states of a trait T , the total number $|R^*|$ of relations in the relative relational diversity \mathcal{D} (as given by equation (1a)) is specified by $|R^*| = \sum_{(i,j) \in R_T} n_{i,j}$, where $n_{i,j} = n_i \cdot (n_j - \delta_{i,j})$ under the TO-perspective. Herein n_i specifies the number of objects showing the i -th trait state (i.e. objects x with $T(x) = i$), R_T is the relation between the trait states, and $n_{i,j}$ is the number of relations $(x, y) \in R$ with $T(x) = i$ and $T(y) = j$.

The transition from R (or R^*) to R_T implies a probably less obvious change in the classical definition of mathematical relations. The mathematical definition is based on the statement of the presence or absence of a relation between two objects, where the objects are not assigned any weight. This is different in R_T , since there the objects are types (trait states) that are associated with frequencies as weights. The presence-absence statement about relations, however, still applies, since types now take the place of objects. Therefore, in the realm of diversity measurement, it is meaningful to extend the classical notion of relations by assigning weights to relations. Hence, in this case the numbers $n_{i,j}$ are considered as weights of relations.

It remains to clarify, how the interpretation of diversity in terms of sampling without replacement enters the weights. In fact, sampling requires units to be sampled, and this gains relevance with respect to weights only if they can be expressed in a commensurable way by numbers of units of measurement. Given this situation, the definition of relative relational diversity given in equation (1a) applies consistently. Otherwise, if weights

are not *a priori* commensurable, the implied continuum of measurements can be represented by infinitely large collections. This situation is most appropriately represented by changing from absolute to relative frequencies, i.e. by making use of $p_i := n_i/n$ and $p_{i,j} := n_{i,j}/(n \cdot (n - 1))$. Hence, equation (1a) becomes

$$\mathcal{D} = 1 - \sum_{(i,j) \in R_T} p_{i,j} \quad (1b)$$

Under the TO-perspective, where $n_{i,j} = n_i \cdot (n_j - \delta_{i,j})$, it follows that $p_{i,j} = \frac{n}{n-1} \cdot p_i \cdot (p_j - (\delta_{i,j}/n))$, so that for infinitely large n one obtains $p_{i,j} = p_i \cdot p_j$. Note that $p_{i,j} = 0$ for $(i, j) \notin R_T$, so that the sum in equation (1b) could as well be extended to all trait states i and j .

Weights other than natural numbers are quite common in diversity studies in ecology and agriculture, where areas occupied by individuals, species or cultivars or their biomasses are of interest. The weights used in such studies are primarily given to types, where relations between types seem to play no role so far. In any case, the appropriate perspective is the TO-perspective, so that the weights w_i , say, replace the numbers n_i . These weights have of course a dimension, which does however not affect the measurement of diversity, since only the relative quantities $p_i = w_i / \sum_j w_j$ enter the computation of \mathcal{D} . The intrinsically quantitative nature of weights entails collections of infinite size so that in equation (1b), $p_{i,j}$ has to be replaced by $p_i \cdot p_j$ with the result that $\mathcal{D} = 1 - \sum_{(i,j) \in R_T} p_i \cdot p_j$.

Explicit measure of diversity – effective number

“The effective number ... is at the core of the concept of diversity in biology” (Jost 2006), where effective numbers are understood to directly refer to numbers of different types. In many if not most situations, measures of diversity are conceived in implicit terms such as the above measure \mathcal{D} of relative diversity. However, by virtue of the concept of diversity, the implicit statement is incomplete unless it can be consistently explicated in terms of numbers of types. Turning an implicit measure of diversity into an explicit one (as specified by the effective number), therefore requires a conceptual underpinning of the involved transformation. Herewith the focus is set on an “ideal model”, in which the implicit measure is unambiguously defined by the number of types. Equating the observed measure with that obtained from the ideal model and solving for the number of types provides the desired transformation into an explicit measure in terms of the “effective” number of types. The adjective “effective” is therefore directly bound to the chosen ideal model (see Gregorius 1991).

The generally agreed ideal model in diversity measurement is specified by a uniform distribution of types and thus by a collection, in which all types are equally represented. Only in such a model of complete evenness the number of types is unambiguously defined (see e.g. McArthur 1965). This model can be expressed in terms of a special relation known as *equivalence relation* (see e.g. Podani, 2006). Each equivalence relation implies a decomposition of the collection into disjoint subsets (i.e. equivalence classes) such that each member of a class is related to every other member of that class but is unrelated

to all objects outside the class. Each object can therefore be uniquely assigned to a type that is defined by the equivalence class to which the object belongs.

The assignment establishes a trait with classes as trait states (or types). The representation of types is then specified by the number, proportion or weights of objects in the corresponding class. This approach is compelling, since it makes use of the fact that each trait can be represented by an equivalence relation and vice versa. At a last step, the equivalence relation becomes the ideal model by requiring equal representation for all equivalence classes.

In this case, the number $|C_K| - |R^*|$ of unrelated pairs of objects equals $k \cdot (n - k) \cdot v$, where v is the number of classes (types) and k is the number of members of each class. Equation (1a) therefore becomes

$$\mathcal{D} = \frac{k \cdot (n - k) \cdot v}{n \cdot (n - 1)}$$

Since $k \cdot v = n$, so that $k = n/v$, this can be written as

$$\mathcal{D} = \frac{n}{n - 1} \cdot \left(1 - \frac{1}{v}\right)$$

According to the above explanations, the transformation of relative into explicit diversity is now achieved by solving the last equation for v :

$$v = \left(1 - \frac{n - 1}{n} \cdot \mathcal{D}\right)^{-1} \quad (2)$$

In equation (2), \mathcal{D} can be chosen to equal any observed value of the relative relational diversity in a collection of n objects, which then turns v into an effective number based on an equivalence relation with equally sized equivalence classes as the ideal model. This justifies introduction of v as a measure of *explicit relational diversity* (for an illustration see Figure 1).

Apparently, $v \leq n$, which is in accordance with the expectation that explicit diversity should never exceed the size of the base collection. The fact that $\mathcal{D} = 1$ implies $v = n$ gains special significance if one considers the classical species count. In this situation, the objects of interest are not the individual members of the respective species but rather the species themselves. Hence, the assessment of the species richness (diversity) of a community builds on a collection made of species, so that all objects in such a collection differ from each other and are therefore unrelated. The effective number thus equals the collection size. This example may serve to demonstrate the necessity of a clear and meaningful specification of a collection and its objects.

Another example comes from the situation, where the objects are types that are assigned weights. These weights replace the type frequencies. As was explained earlier, this situation can be taken into account by letting collection sizes approach infinity so that equation (2) becomes $v = (1 - \mathcal{D})^{-1}$. Thus by equation (1b), $v = \left(\sum_{(i,j) \in R_T} p_i \cdot p_j\right)^{-1}$. If no relations exist between different types (so that $(i, j) \notin R_T$ if $i \neq j$), v reduces to

the well known effective number $v = (\sum_i p_i^2)^{-1}$ (in a population genetic context see e.g. Crow & Kimura 1970, p.323-324).

Diversity at several levels – community diversity

A problem may arise with the concept of relational diversity if, for example, the members of a biological community under consideration belong to several systematic categories. In such communities the diversity could be expected to increase with the number of individuals belonging to different categories. Thus, for 100 individuals belonging to different species of the same genus the diversity is rated to be lower than for 100 individuals belonging to different genera of the same family, and this diversity is again rated lower than for 100 individuals belonging to different families. However, indecision may exist if the case of two species from different families (say 50 individuals from each species) is to be compared with the case of four species (say each represented by 25 individuals) from three genera of the same family.

The problem is obviously caused by the fact that the diversity of each systematic category is primarily described by the number of its sub-categories and that therefore no measure of diversity is given that spans several and quite different systematic categories co-occurring in the same community. Hence, in such situations the relations among the members of a community must include characteristics of multiple systematic categories in order to enable comparison of different communities with respect to their diversity. The probably most obvious method of designing such relations is provided by a measure of difference between individuals that reflects the distinctiveness of the categories. Taxonomic distance would be such a measure (see e.g. Weikard et al., 2006). More generally, any measure would be appropriate that yields smaller differences within than between categories.

As was mentioned above, given a proper measure of difference between objects, a relation can be defined for each fixed threshold of difference by declaring two objects as related if their difference does not exceed the threshold. In this way, by choosing variable thresholds, a family of relations is generated, for each of which a measure of diversity is defined by equations (1) and (2). The fact that two objects related for a particular threshold are necessarily related for any higher threshold difference implies a hierarchical ordering of the relations. Thus, if relation R' is generated by a lower threshold value than relation R'' , then $R' \subseteq R''$ and consequently the diversity associated with R'' cannot exceed the diversity associated with R' . In other words, the diversity of a collection decreases (not necessarily strictly) with increasing threshold value. Hence, in the example of systematic categories, different communities can be compared for their diversity values only when applying the same measure of difference and threshold value. It may then turn out, that the superiority in diversity of one community over another may depend on the chosen threshold level.

The nestedness of relations implied by increasing threshold values can be interpreted in a more comprehensive framework. Since unrelatedness indicates distinctiveness, an increasing number of unrelated objects can be expected from the application of more

efficient methods of distinguishing objects. Hence, a method that detects more unrelated objects can be awarded a higher resolving capacity. In more formal terms this states that for two nested relations R' and R'' with $R' \subset R''$, R' is more distinctive than R'' and thus realizes a higher “resolution”. The obvious fact that methods of higher resolution reveal higher diversity in a collection can thus be traced back to the underlying relations.

Order relations and hierarchical relations

Many forms of biological interaction are characterized by asymmetric relations such as can be found in food webs, where one species depends directly on another for its diet but not vice versa. Dominance relationships among the species in a community, among members of a population, or among alleles at a gene locus can be mentioned as other examples. These relations constitute *strict order relations* in the sense that $(x, y) \in R$ implies $(y, x) \notin R$ (asymmetry), and $(x, y) \in R$ and $(y, z) \in R$ implies $(x, z) \in R$ (transitivity). They seem to have attracted little attention, if any, in quantitative analyses of biological diversity.

The same seems to be true for another type of relation that implies some kind of order and plays an important role in biology. This type is realized by hierarchically organized interactions between objects. Hierarchical organization avoids conflicts that may arise from competing dominance relations as they would occur if two objects would jointly and independently dominate a third one. Joint dominance of two objects over a third is therefore not allowed “in parallel” but only “in succession”, i.e. if one of the two objects already dominates the other. The interactions within most biological communities as well as the interactions along food chains are more or less strongly hierarchically organized (for an overview with emphasis on hierarchies of community organization see Sugihara et al. 2003). Herewith, the dominance relationships need not be realized directly between two objects but may rather require a chain of sequential direct dominance relations. This situation will be referred to by the wording “one object is *chain-related* to another”.

For a formal representation of a *hierarchical relation* that excludes parallel actions in the above sense it is thus required that if two objects are both chain-related to a third, then one of the two objects should be chain-related to the other (this is not to be confused with what is occasionally called a Euclidean relation). To put the notion of chain-relatedness on a firmer basis, the *transitive closure* \bar{R} of a binary relation R will be introduced. For this purpose consider an object x as chain-related to an object y within R if there exists a series of objects $z_i \in K$ ($i = 1, \dots, m$) such that $z_1 = x$, $z_m = y$, and $(z_i, z_{i+1}) \in R$ for $i = 1, \dots, m - 1$. Then

$$\bar{R} := \{(x, y) \mid x \in K, y \in K, x \text{ is chain-related to } y \text{ within } R\}$$

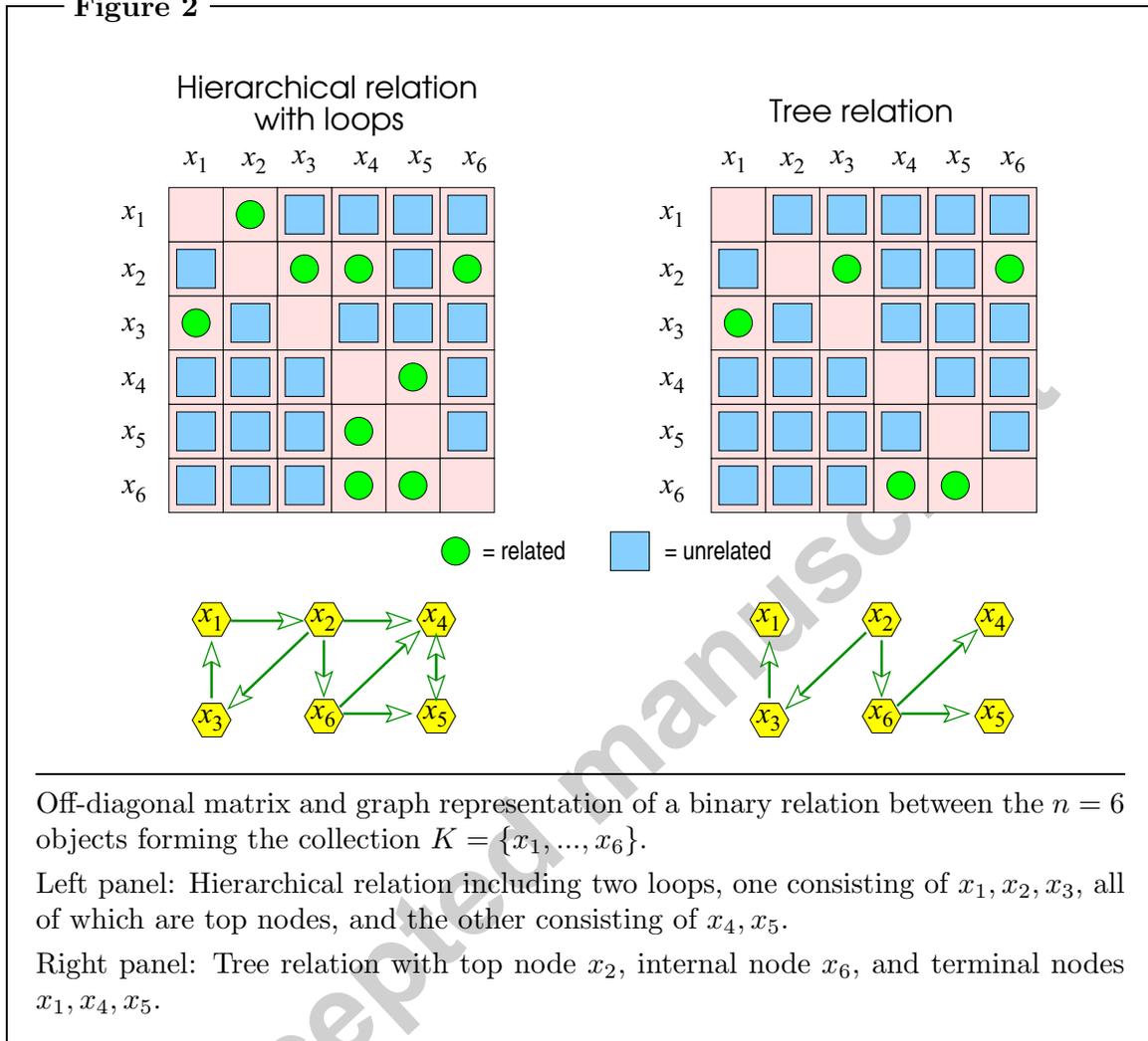
is a transitive relation. The above explanations can then be stated as (see Figure 2 for an illustration)

- ▷ A relation R is hierarchical if $(x, z) \in \bar{R}$ and $(y, z) \in \bar{R}$ implies that for $x \neq y$ either $(x, y) \in \bar{R}$ or $(y, x) \in \bar{R}$.

In essence, by introducing \bar{R} , this definition enforces transitivity in cases where the relation R is not yet transitive. Therefore, if one requires transitivity in addition to

the hierarchy condition (so that $\bar{R} = R$), the above definition remains the same if \bar{R} is replaced by R . In this case the relation will be called *strictly hierarchical*.

Figure 2



To demonstrate more explicitly the hierarchical set structure associated with a hierarchical relation, consider for each object x its *dr-class* as the set $R_{dr}(x)$ of all objects to which it is (directly) related, i.e.

$$R_{dr}(x) := \{y \in K \mid (x, y) \in R\}.$$

Continuing the previous approach, the extension of the dr-class from directly related to chain-related objects gives rise to the definition of the *cr-class* $R_{cr}(x)$ of an object x by the set of all objects to which x is chain related, i.e.

$$R_{cr}(x) := \{y \in K \mid (x, y) \in \bar{R}\}.$$

Note that, even for a transitive relation, $R_{dr}(x)$ differs from $R_{cr}(x)$ if $R_{dr}(x)$ contains at least one object that is related to another object.

It now follows from the transitivity of the relation \bar{R} that $y \in R_{cr}(x)$ implies $R_{cr}(y) \subseteq R_{cr}(x)$. To see this, recall that for $y \in R_{cr}(x)$ and $z \in R_{cr}(y)$ one obtains $(x, y) \in \bar{R}$ and $(y, z) \in \bar{R}$, so that $(x, z) \in \bar{R}$ and thus $z \in R_{cr}(x)$. Since this holds for

any $z \in R_{cr}(y)$ one obtains the desired result $R_{cr}(y) \subseteq R_{cr}(x)$. Moreover, from the hierarchy condition it follows that the intersection of two cr-classes is either empty or one is contained in the other. To see this suppose that $z \in R_{cr}(x) \cap R_{cr}(y)$. Then $(x, z) \in \bar{R}$ and $(y, z) \in \bar{R}$ so that either $(x, y) \in \bar{R}$ or $(y, x) \in \bar{R}$. Without loss of generality, let $(x, y) \in \bar{R}$; then $y \in R_{cr}(x)$ and, by transitivity of \bar{R} , $R_{cr}(y) \subseteq R_{cr}(x)$. This proves the assertion. The same steps of proof show that for a strictly hierarchical relation either $R_{dr}(y) \subseteq R_{dr}(x)$ or $R_{dr}(y) \cap R_{dr}(x) = \emptyset$.

Hence, the cr-classes of a hierarchical relation establish a collection of nested subsets that directly reflect our notion of a hierarchical structure. The same holds true for the dr-classes of a strictly hierarchical relation. The notion is also associated with the common illustration by a tree-like graph, in which an object appears as an “internal node” if it is simply related to at least two other objects that are not chain-related to each other. Herewith, object x is called *simply related* to object y , if $(x, y) \in R$ and no other chain relates x to y . Top nodes are internal nodes whose cr-classes are not contained in the cr-class of any other object. This implies, that no other node is related to a top node. At the other extreme, the cr-class of an object may be empty or consist only of the object itself, in which case this object is frequently called a “terminal node” of the tree.

However, the above definition of a hierarchical relation is not yet sufficiently specific to reflect the common notion of a tree structure. In fact hierarchical relations do not even exclude the possibility of loops in chains. A loop chain-relates an object to itself so that $(x, x) \in \bar{R}$ for any such object. This implies in particular that $x \in R_{cr}(x)$ only if x is part of a loop (see Figure 2 for an illustration). The ubiquity of feedback mechanisms at various levels of organization in biological systems (extending from biochemical feedback regulation over reciprocal altruism to nutrient cycling and general hypercycles) can probably be summarized under the category of hierarchical relations in which all objects belong to loops. Their diversity is well-documented but hardly (if ever) quantified.

To distinguish these kinds of structure from a typical tree structure, one apparently requires at least the absence of loops. Moreover, since the branches of trees diverge but never converge, no two different objects are allowed to be jointly related to a third object (i.e. if $(x, z) \in R$ and $(y, z) \in R$ then $x = y$). This condition will be referred to as *divergence*. However, if the condition of divergence applies, then situations for which the hierarchy or strict hierarchy condition become effective do not occur. It is therefore sufficient to define a *tree relation* as a divergent relation in which no loops occur. Tree relations are thus asymmetric and hierarchical. It should be noted that the graphs associated with tree relations are not the same as the graphs commonly addressed as a tree in graph theory (see Figure 2 for an illustration).

In classical cladograms or phylogenies as examples of tree relations, only the terminal nodes are observable, and internal nodes are fictitious objects that result from the method of dendrogram construction (or reconstruction of descent). Observed and fictitious objects together form the collection on which the tree relation is defined, where the relation itself is completely specified by the dendrogram. Given such a dendrogram, a tree relation between the objects (nodes) can be defined by the existence of a connecting branch,

where the direction of the branch always points towards a terminal node (thus node x is related to node y if x is connected to y by a branch along which y is closer to a terminal node than x). In this way the number of relations is unambiguously determined, and the diversity measure \mathcal{D} is well-defined. With this definition of a relation between the nodes of a dendrogram it can be shown that for any given number of terminal and internal nodes the number of relations tends to decrease with increasing number of disjoint groups. Thus, \mathcal{D} reflects topological properties of the dendrogram. By referring to a number of classes or groups, the effective number v is particularly intuitive in this case as a measure of topological diversity.

Normalization problems – constraints

Different types of relation may imply different constraints as to the maximum numbers of relations c_{max} that they allow for and the minimum number c_{min} they enforce. This could be taken as an argument for adjusting the $|C_K|$ in equation (1a) because the maximum number c_{max} of admissible relations is smaller than $n \cdot (n - 1)$. Similarly, the minimum number, which is always zero in equation (1a), may become positive under certain constraints. Taking these constraints into account, the measure \mathcal{D} could be stated in the more general form $\mathcal{D} = (c_{max} - |R^*|)/(c_{max} - c_{min})$.

The measurement of diversity in strict order, in hierarchical or in tree relations can be drawn on as examples of constraints. Since a strict order relation excludes symmetrical relations, the maximum number of relations between n different objects is halved under this constraint, i.e. $c_{max} = \frac{1}{2}n(n - 1)$. At the other extreme, all objects are allowed to be unrelated, which implies $c_{min} = 0$. Complete homogeneity or the absence of structure, i.e. $\mathcal{D} = 0$, would then be realized, if for any two different objects x and y either $(x, y) \in R$ or $(y, x) \in R$. This implies that the strict order is complete (or linear), and all objects are aligned in one chain. Maximum diversity, i.e. $\mathcal{D} = 1$, again corresponds to the absence of any relations.

A particularly strong reduction of c_{max} results from the constraints implied by mating relations that are defined by the fusion of gametes. No object (gamete) can realize more than one relation to other objects, and the relation is symmetric. The totality of n objects is divided into two sets of size n' and n'' , say, which refer to the two gametic sexes. The number of potential relations is thus restricted to $c_{max} = \min\{n', n''\}$. When the number of male and female gametes are not considered as constraints, c_{max} is given by the largest natural number not exceeding $n/2$.

Another example for constraints that imply $c_{min} > 0$ is given by an asymmetric relation in which exactly m of a total of n objects are not related to other objects and each of the remaining $n - m$ objects is related to at least one other object. Tree relations between n objects with m terminal nodes fulfill these constraints in particular. The minimum number of relations is of course realized if each of the $n - m$ objects is related to exactly one other object so that $c_{min} = n - m$. The maximum number of relations results if all admissible relations among the $n - m$ occur, which are $\frac{1}{2}(n - m) \cdot (n - m - 1)$ relations, and if each of the $n - m$ objects is related to each of the m objects, which yields $(n - m) \cdot m$

relations. Hence, $c_{max} = \frac{1}{2}(n - m) \cdot (n - m - 1) + (n - m) \cdot m = \frac{1}{2}(n - m) \cdot (n + m - 1)$.

The conversion of these relative measures of diversity to an explicit measure by equation (2) via the ideal model is unaffected by the various normalizations. It has however to be taken into consideration that constraints that are incorporated into the measurement of diversity ignore the remaining potential relations as sources of diversity, so that the reduction in diversity due to the constraints remains unattested. For example, it might be of interest to compare the effects of different kinds of organization of biological systems on diversity. Candidates could be the constraints imposed by hierarchical as compared to those imposed by special kinds of reticulate organization. Such comparisons can be problematic, when only the accordingly normalized versions of relative diversity are used. It could therefore be preferable to make use of both the normalized and non-normalized versions of relative diversity.

Equivalence relations as ideal model for the determination of an effective number might be argued to have no intuitively obvious connection to certain types of relation like the hierarchical, and that therefore the ideal model should be adapted to the constraints. Such objections can be countered with the help of the cr-classes or dr-classes. These classes are defined for all relations and specify sets of objects that reflect the constraints of the respective relation. Equivalence classes are then simply the dr-classes generated by equivalence relations (cr-classes are identical to dr-classes for equivalence relations). It is therefore meaningful to generally address the explicit measure v of diversity as given by equation (2) as an *effective number of classes*.

Extension to include samples of size m

So far the assessment of diversity was based on counting binary relations between objects. It was pointed out that this is equivalent to sampling two objects one after the other (without replacement) and state whether the first is related to the second. Retaining the implied concept of assessing diversity on the basis of unrelatedness found in samples taken without replacement, one may extend the sample to sequences of more than two different objects and again state whether each is related to its successor in the sequence. Resuming the above idea of chain-relatedness, this gives rise to the consideration of sequences of given length m , say, in which each object is related to its successor. Such a sequence will be called a *chain* (of length m). A chain starting with object x and ending with object y thus implies that y belongs to the cr-class of x , i.e. $y \in R_{cr}(x)$.

Diversity will then be evaluated with reference to the number of chains of length m , which the relation R allows to form, in comparison with the total number of the sequences that can be obtained by sampling m objects without replacement from the collection. The most characteristic consequence of increasing the sample size consists in an increase of the chances to recover unrelated objects. In fact, if unrelatedness has not been found in a sample of $m - 1$ objects, there will still be a chance to find it by insertion of another of the remaining objects somewhere into the sample sequence. Hence, to arrive at the same assessment of diversity for smaller samples, the number of relations must be reduced. Moreover, long chains can be formed only from larger sets of interrelated objects. With

increasing m , the evaluation of diversity therefore depends to an ever higher degree on only these sets. In the case of equivalence relations these sets are given by the largest classes.

Given a collection of n objects ($n \geq m$), there are $[n, m] := n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - m + 1)$ sample sequences each consisting of m different objects. The set of these sequences will be denoted by C_K^m (with $C_K^2 = C_K$) and among these the set of chains of length m will be referred to as R_m^* (with $R_2^* = R^*$). For convenience, the notation $r_m := |R_m^*|$ will be used in the following (also recall that $[n, m] = |C_K^m|$). The previous definition of a measure of diversity given by equation (1a) (where m was equal to 2) can now be extended to arbitrary values of $m \geq 2$ through replacement of $|C_K|$ and $|R^*|$ by $[n, m]$ and r_m , respectively:

$$\mathcal{D}_m = \frac{[n, m] - r_m}{[n, m]} \quad (3)$$

In terms of probabilities, \mathcal{D}_m equals *the probability of sampling without replacement a sequence of m objects ($m \leq n$) of which at least one is not related to the object following it in the sequence*. By definition, the probability becomes 1 only if $r_m = 0$ and thus if m exceeds the maximum chain length. Moreover, as was argued above, the probability of encountering unrelated objects increases (and so does \mathcal{D}_m) with the sample size m .

To obtain the effective number recall that for equivalence relations chains are made of objects belonging to the same equivalence class. Thus, if the i -th class has size n_i , there are $[n_i, m] := n_i \cdot (n_i - 1) \cdot (n_i - 2) \cdot \dots \cdot (n_i - m + 1)$ possibilities to compose chains of length m from that class. For $n_i < m$ it will be understood that $[n_i, m] = 0$. Given s classes ($\sum_{i=1}^s n_i = n$), the total number of chains thus amounts to $r_m = \sum_{i=1}^s [n_i, m]$, and one arrives at

$$\mathcal{D}_m = 1 - \frac{\sum_{i=1}^s [n_i, m]}{[n, m]} \quad (4)$$

For the ideal model with v equivalence classes, $n_i = k$ so that $v \cdot k = n$, and by insertion into equation (4) and setting the result equal to equation (3) one obtains $r_m = v \cdot [n/v, m]$ as the equation to be solved for v (note that $[n, m]$ is well defined for natural numbers m and general real numbers n , and $[n, m] \geq 0$ for $n \geq m - 1$). Since $v \cdot [n/v, m] = n \cdot [n/v - 1, m - 1]$, this equation can be rewritten as

$$\frac{r_m}{n} = \left[\frac{n}{v} - 1, m - 1 \right] \quad (5)$$

The equation has a unique solution $v = v_m$, because its right side is a strictly decreasing function of v over its relevant values; the solution can be obtained by simple numerical iteration. The relevant range for v is determined by the fact that $[n/v - 1, m - 1] \geq 0$ for $n/v - 1 \geq m - 2$, and since $r_m \leq [n, m]$. The effective number v_m thus always lies in the interval $1 \leq v_m \leq n/(m - 1)$.

For *infinitely large collections* the effective number can be obtained with the help of equation (4) by letting n tend to infinity while retaining the proportions $p_i = n_i/n$ of the equivalence classes (as was previously done to arrive at equation (1b)). Writing

$$\frac{[n_i, m]}{[n, m]} = \prod_{j=0}^{m-1} \frac{n_i - j}{n - j} = \prod_{j=0}^{m-1} \frac{p_i - \frac{j}{n}}{1 - \frac{j}{n}}$$

and passing to the limit $n \rightarrow \infty$ one obtains $[n_i, m]/[n, m] \rightarrow p_i^m$ and therefore

$$\mathcal{D}_m = 1 - \sum_{i=1}^s p_i^m$$

Note that this representation does not only apply to the equivalence classes but is more generally true under the TO-perspective, when no relations exist between types (indexed by i) and all objects of the same type are mutually related.

The ideal model, when applied to the case of infinite collections, requires $p_i = 1/v$ for $i = 1, \dots, v$, where v is the number of classes in the ideal model. Solving the resulting equation $\mathcal{D}_m = 1 - v^{1-m}$ for v yields as the effective number

$$v_m = (1 - \mathcal{D}_m)^{\frac{1}{1-m}} \quad (6)$$

Under the above-mentioned conditions where $\mathcal{D}_m = 1 - \sum_{i=1}^s p_i^m$, equation (6) assumes the well-known form

$$v_m = \left(\sum_{i=1}^s p_i^m \right)^{\frac{1}{1-m}}$$

The family v_m of diversity measures was suggested by Hill (1973) for arbitrary real values of m . Interestingly, Patil & Taillie (1982) arrive at the same effective number even though their considerations proceeded from an apparently different concept of implicit diversity based on the average of a particular measure of rarity of types.

Further analyses particularly of the effect of m on the measure v_m can be found in Hill (1973) or Gregorius (1978). These analyses confirm the above general characterization of the dependence of diversity measurement on larger sets of interrelated objects. In particular, v_m decreases with increasing m and ultimately depends only on the largest of the frequencies p_i . Moreover, v_0 specifies the number of types (richness) and v_1 results from the limit of v_m as m approaches 1. It should, however, be kept in mind that equation (6) has only been shown to have an interpretation in the context of binary relations on infinite collections and natural numbers $m \geq 2$.

Concluding remarks

The concept of diversity differs intrinsically from the concept of dispersion in that it is aimed at providing information on the variety and multiformity of collections and thus on numbers of distinguishable objects. It therefore relies on discretizable and countable units of perception. The elementary principle of counting types becomes problematic when they are represented in different amounts and are distinguishable to different degrees. It is shown in this paper that these problems can be overcome by switching from counting types to counting pairs of objects that are characterized in a binary fashion as being related or not related. Counting pairs is extendable to counting sequences consisting of more than two objects. Relatedness can be defined in innumerable ways, some categories of which were addressed above. The concept of effective number in turn allows switching

back again from counting relations to counting types as the true basis for measuring diversity. This principle can be applied to a wide range of problems and leads in each case to an explicit and interpretable measure of diversity.

Particular community ecologists may miss reference to the Shannon index in the above demonstrations. As is well known, this index is contained in the family v_m via the transformation $\log v_1$, under the premise that m equals any non-negative real value. Initially, the Shannon index was derived as a measure of information content and, in physics, as a measure of entropy. Neither interpretation has a direct bearing on the notion of diversity as an effective number. It may therefore be concluded that the Shannon index is not part of the present concept of relational diversity or, more extremely, that it is not a measure of diversity at all. However, the interpretation of the family v_m on the basis of chains or samples of relations may not be the only way to arrive at this family. Interpretations in terms of relations based on other sampling procedures and transformations may justify extension of the family v_m to arbitrary real values of m .

The need for indices that go beyond mere counts of types (mostly referred to as “richness”) is usually felt when diversity is to be assessed for more complex features. One characteristic of complexity that is widely agreed upon is seen in the multiplicity of relations among the parts of a system. Relations are thus the basic determinants of complexity, and any quantification of the degree of complexity is therefore directly concerned with the measurement of relational diversity. Application of the present concept to the assessment of complexity will be an interesting task.

Acknowledgement: The suggestions of an anonymous reviewer helped to clarify the demonstrations about objects and types.

References

- Crow JF, Kimura M 1970. *An Introduction to Population Genetics Theory*. Harper & Row, New York, Evanston, London
- Gregorius H-R 1978. The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. *Math. Biosci.* 41: 253-271
- Gregorius H-R 1991. On the concept of effective number. *Theoretical Population Biology* 40: 269-283
- Gregorius H-R, Gillet EM 2008. Generalized Simpson-diversity. *Ecological Modelling* 211: 90-96
- Hill MO 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427-432
- Jost L 2006. Entropy and diversity. *OIKOS* 113: 363-375
- MacArthur RH 1965. Patterns of species diversity. *Biol. Rev.* 40: 510-533
- Myers W, Patil GP 2006. Biodiversity in the age of ecological indicators. *Acta Biotheoretica* 54: 61-160
- Patil GP, Taillie C 1982. Diversity as a concept and its measurement. *Journal of the American Statistical Association* 77(379): 548-561

- Podani J 2006. With a machete through the jungle: some thoughts on community diversity. *Acta Biotheoretica* 54: 125-131
- Rao R 1982. Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* 21: 24-43
- Ricotta C 2006. Towards a complex, plural and dynamic approach to diversity: rejoinder to Myers and Patil, Podani, and Sarkar. *Acta Biotheoretica* 54: 141-146
- Sarkar S 2006. Ecological diversity and biodiversity as concepts for conservation planning: comments on Ricotta. *Acta Biotheoretica* 54: 133-140
- Shimatani K 2001. On the measurement of species diversity incorporating species differences. *Oikos* 93: 135-147
- Staudhammer CL, LeMay VM 2001. Introduction and evaluation of possible indices of stand structural diversity. *Can. J. For. Res.* 31: 1105-1115
- Sugihara G, Bersier L-F, Southwood TRE, Pimm SL, May RM 2003. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl. Acad. Sci. USA* 100(9): 5246-5251
- Weikard H-P, Punt M, Wesseler J 2006. Diversity measurement combining relative abundances and taxonomic distinctiveness of species. *Diversity and Distributions* 12: 215-217
- Weitzman ML 1992. On diversity. *The Quarterly Journal of Economics* 107(2): 363-405
- Wootton JT 2005. Field parameterization and experimental test of the neutral theory of biodiversity. *Nature* 433: 309-312
- Xu Y, Olman V, Xu D 2002. Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning trees. *Bioinformatics* 18: 536-545