Concavity and additivity in diversity measurement: re-discovery of an unknown concept

Sönke Hoffmann

Faculty of Economics and Management (VWL III), Otto-von-Guericke University Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany

Abstract

Social and natural sciences employ a number of different measures of distribution based diversity. This paper presents a unifying (two-parameter) notation for most of these, which is derived from three different one-parameter generalizations of the Shannon entropy. The model enables scientists and decision makers to measure distribution based diversity in a new, more flexible manner, and represents a useful complement to models of generalized feature based diversity, such as Nehring and Puppe's (2002) "Theory of Diversity". It is shown by example how trade-offs between important diversity properties can be made explicit within such general framework. Although more than thirty years old, the model seems to be entirely unknown in economics and ecology.

Key words: Diversity measurement; Generalization; Additivity; Concavity
JEL classification: Q20, C65

1 Introduction

Diversity is a very diverse issue. Ideas of how to measure it vary as much as the contexts to which they are applied. The most general classification of diversity measures divides them into two groups: *Qualitative* diversity is defined as the degree of differences among diversity units (often: products, choices, DNA's or molecules), whereas *distribution based* diversity is determined by the abundance distribution among diversity units (often: species, financial assets or households) [1]. The pioneering work of Weitzman (1992) on

Email address: soenke.hoffmann@ww.uni-magdeburg.de (Sönke Hoffmann).

[1] See Baumgärtner (2006) for a conceptual overview.

qualitative diversity was followed by profound generalizations such as Nehring and Puppe's (2002) "Theory of Diversity". Distribution based diversity, on the other hand, still lacks a single model, which includes all frequently used measures of this class. At least two fundamentally different generalized models of distribution based diversity coexisted during the last decades. The so called "Rényi diversity" is an additive and non-concave one-parameter generalization of the Shannon entropy (Rényi, 1961). It is successfully applied to biodiversity measurement (e.g. Ricotta 2003), statistical sampling (Mayoral, 1998), economic diversity modelling (Beran, 1999), and many other contexts. A second generalization of Shannon entropy, which is concave but not additive, was independently derived by information theorists Havrda and Charvát (1967), statisticians Patil and Taillie (1982) and physicist Tsallis (1988) and recently re-discovered for biodiversity measurement by Keylock (2005) [2].

We present a unifying two-parameter notation, which includes additive *and* concave generalizations of Shannon entropy. The Shannon entropy itself, as well as most other special indices of distribution based diversity can be simply recovered by a corresponding pair of parameter values. These include indices of Rényi diversity, Tsallis diversity, an unknown "Gaussian" one-parameter diversity class and the well-known "generalized effective number", being most often used to measure biological diversity (Hill, 1973), industry concentration (Hannah and Kay, 1977) or fragmentation (Laakso and Taagepera, 1979).

Some authors find the Rényi class admissible (Ricotta, 2003), some may prefer the Tsallis model Keylock (2005) and others even regard the effective number as the only "true" diversity Jost (2006). However, by accepting diversity *per se* as a multi-faceted phenomenon that can be defined and measured in many different ways, conflicts between different ideologies on the measurement of distribution based diversity could quickly be resolved. The model presented enables scientists and decision makers to measure distribution based diversity in a much more flexible and context-driven manner because it provides a direct link between underlying properties of the measurement concepts and two simple parameters. To illustrate that, we derive the parameter spaces for additivity and concavity. Due to its general character the model is very complementary to generalized models of qualitative diversity and, therefore, fills a large gap in diversity measurement. Surprisingly, this "new" concept exists in information theory, ready to use, for more than thirty years but it is still unknown in most social and natural sciences.

[2] This formalism is most widely spread under the name 'Tsallis' entropy although Havrda and Charvát (1967) have derived it more than two decades earlier. Similarly, the 'Shannon' entropy is not due to the father of information theory C.E. Shannon but due to mathematician von Neumann (1927). However, in this paper we will refer to the most popular nomenclature. The 'true' origin of a mathematical expression is most often too hard to recover.

The remainder is organized as follows. Section 2 summarizes some mathematical and information theoretical background. In the preceding sections 3 to 5 we briefly review three different ways to generalize Shannon entropy, including the approaches of Rényi and Tsallis. The Sharma-Mittal formalism is introduced in section 6 and additivity and concavity conditions are summarized for its most important special cases. Finally, sections 7 and 8 conclude and critically give an outlook. The appendix includes some hypothetical example data, the formal representation of additivity and concavity properties and proofs.

2 Prerequisite

2.1 Distribution based diversity

Distribution based diversity only depends on the given number of classes n and the way individuals or abundances are distributed over these n classes. Let $\mathcal{P}$ be the n dimensional unit simplex and $\mathbf{p} \in \mathcal{P}$ then $V : \mathbf{p} \rightarrow \mathbb{R}^+$ denotes a (distribution based) diversity function. Throughout this paper superscript letters of diversity functions will indicate the names typically connected with them, subscripts represent parameters of that function and/or the index $i = 1 \ldots n$. The most fundamental assumption on V is that, if all individuals are equally distributed, the set of classes is maximally diversified and if all individuals are unified in a single class the set of classes is minimally diversified[3]. In other words, let $\check{\mathbf{p}}$ be a completely uneven distribution of abundances (one $p_i = 1$ and all other $p_{k \neq i} = 0$) and $\bar{\mathbf{p}}$ a completely even distribution (all $p_i = \frac{1}{n}$) then the inequality

$$0 \leq V(\check{\mathbf{p}}) \leq V(\mathbf{p}) \leq V(\bar{\mathbf{p}}) \tag{1}$$

should hold.

2.2 Additivity and concavity

Let $A = \{A_i\}_{i=1 \ldots n}$ and $B = \{B_j\}_{j=1 \ldots m}$ be two non-overlapping classifications on a set of individuals with $\mathbf{p}_i = \{p_i\}_{i=1 \ldots n} \in \mathcal{P}$ and $\mathbf{p}_j = \{p_j\}_{j=1 \ldots m} \in \mathcal{P}$ denoting the relative abundances of A and B.

[3] In other disciplines this diversity property is also known as 'Maximum Value Principle' (Hilderman and Hamilton, 2001). See Rao (1982) for a set of axioms that characterize distribution based diversity.

Property 1 (Additivity) *V is called additive if*

$$V(\mathbf{p}_{ij}) = V(\mathbf{p}_i) + V(\mathbf{p}_j) \qquad (2)$$
$$\textit{viz. } V(A \cap B) = V(A) + V(B).$$

Property 2 (Non-additivity of degree b) *V is called non-additive of degree b if*

$$V(\mathbf{p}_{ij}) = V(\mathbf{p}_i) + V(\mathbf{p}_j) + (1-b)V(\mathbf{p}_i)V(\mathbf{p}_j) \qquad (3)$$

Note that additivity is equivalent to non-additivity of degree 1.

Property 3 (Concavity) *Let $\mathcal{O}_1 \subset \mathbb{R}^n$ be the positive orthant in $\mathbb{R}^n$ and $\mathcal{C} \subset \mathcal{O}_1$ the set of convex sets on $\mathcal{O}_1$. Two synonymous concavity criterions will be used:*

- Criterion 1: *For all $\mathbf{p}_i, \mathbf{p}_j \in \mathcal{C}$, $m = n$ and $\lambda \in [0, 1]$ measure V is (Jensen-) concave if*
$$(1-\lambda)V(\mathbf{p}_i) + \lambda v(\mathbf{p}_j) \leq V((1-\lambda)\mathbf{p}_i + \lambda \mathbf{p}_j). \qquad (4)$$
- Criterion 2: *For all $\mathbf{p} \in \mathcal{C}$ measure V is (Jensen-) concave if*

$$\textit{Hessian } \mathbf{H}^V \textit{ is negative semidefinite.} \qquad (5)$$

A sufficient condition for $\mathbf{H}^V$ to be negative semidefinite is $\mathbf{p}(-\mathbf{H}^V)\mathbf{p}' \geq 0$ (Debreu, 1952). Note, that strict inequality in (4) implies strict concavity of V on $\mathcal{C}$ and V is strictly concave if $\mathbf{p}(-\mathbf{H}^V)\mathbf{p}' > 0$. Further note, that if V is concave on $\mathcal{C}$ then it is also concave on $\mathcal{P} \subset \mathcal{C}$.

2.3 Shannon diversity

Trying to find a diversity measure that satisfies (1), we can get in line with Patil and Taillie (1982), who interpret distribution based diversity as the *average rarity* of given classes. The rarity of class i is characterized by two simple properties.

Theorem 4 *Any non-negative and additive rarity function of a single class takes the form*

$$v_i := c \log_2 \left(\frac{1}{p_i} \right), \;\; c \in \mathbb{R}^+ \qquad (6)$$

A species having no individual ($p_i = 0$) can be considered infinitely rare (compared to the other species), a species having all available individuals ($p_i = 1$) can be considered not rare at all (compared to the other species). Formally, the linear average of rarity (6) is identical to the Shannon entropy (Shannon,

1948) [4] :

$$V^S(\mathbf{p}) := \sum_{i=1}^{n} p_i \ln\left(\frac{1}{p_i}\right) \tag{7}$$

In the following, we keep up the formal analogy between entropy measures and diversity measures (average rarities) of kind V and refer to (7) as *Shannon diversity.*

Proposition 5 $V^S(\mathbf{p})$ *is additive.*

Proposition 6 $V^S(\mathbf{p})$ *is strictly concave.*

2.4 Quasiliear means

Kolmogorov (1930) and Nagumo (1930) independently derived the most general form of a mean that is still compatible with the well-known Kolmogorov axioms of probability.

Definition 7 (Quasilinear mean) *Let x be a random variable and let ϕ denote a continuous and strictly monotone function, referred to as Kolmogorov-Nagumo function (KN function). If ϕ^{-1} is the inverse of ϕ then*

$$\langle x \rangle_\phi := \phi^{-1}\left(\sum_{i=1}^{n} p_i \phi(x_i)\right) \tag{8}$$

is called the quasilinear mean of x with respect to ϕ.

Hardy et al. (1934) prove that $\langle \cdot \rangle_\phi$ is unique for given ϕ but not vice versa.

Theorem 8 (Invariance of quasilinear means) *Let ϕ_1 und ϕ_2 be KN functions and c_1, c_2 constants, then $\phi_1 = c_1\phi_2 + c_2$, $c_1 \neq 0 \Rightarrow \langle \cdot \rangle_{\phi_1} = \langle \cdot \rangle_{\phi_2}$.*

2.5 Deformation

Let $\exp(x) \equiv \mathfrak{z}^x$, where $\mathfrak{z}$ represents the base of the regular logarithm. Here, we consider the natural logarithm $\ln(x)$ and set $\mathfrak{z} = e$ (Euler number).

Tsallis (1994) introduced the b-deformed logarithm/exponential

[4] Note, however, that we have to set $c = 1$ in order to measure in 'bits', which is the most common unit in information theory. For measurements apart from information theory it is more convenient to set $c = \ln(2)$ in order to measure in 'naturals' (nats) (cf. Theil 1967).

$$
\begin{aligned}
\ln_b^T(x) \; := \; & \frac{x^{1-b}-1}{1-b} && (9)\\
\text{and } \left(\ln_b^T\right)^{-1}(x) = & \left((1-b)x+1\right)^{\frac{1}{1-b}} =: \exp_b^T(x), && (10)
\end{aligned}
$$

which have the properties

$$
\begin{aligned}
& \lim_{b\to 1} \ln_b^T(x) = \lim_{b\to 1} -x^{(1-b)} \ln(x) = \ln(x) && (11)\\
\text{and } & \lim_{b\to 1} \exp_b^T(x) = e^x. && (12)
\end{aligned}
$$

Based on $\ln(\exp(x)) = \ln_b(\exp_b(x)) = x$ we define the following two deformation functions.

Definition 9 (Exponential and logarithmic deformation) *The continuum of functions*

$$
\begin{aligned}
\tau_b^{\exp}(x) \; := \; & \ln_b^T(\exp(x)) = \frac{\exp((1-b)\,x)-1}{1-b}, \; b \neq 1 && (13)\\
\textit{and } \tau_b^{\log}(x) \; := \; & \ln(\exp_b^T(x)) = \frac{\ln((1-b)\,x+1)}{1-b} && (14)\\
= & \left(\tau_b^{\exp}\right)^{-1}(x), \; b \neq 1, \; b < \frac{1}{x}+1
\end{aligned}
$$

will be called the exponential and logarithmic deformation of degree b respectively.

Lemma 10 (Additivity of b-deformed functions) *Let V be an additive function, then $\tau_b^{\exp}(V)$ is nonadditive of degree b.*

Figure 1 shows the exponential and logarithmic deformations of degree b. In the following we will consider only concave deformations.

Proposition 11 *$\tau_b^{\exp}(x)$ is strictly concave for $b > 1$.*

Proposition 12 *$\tau_b^{\log}(x)$ is strictly concave for $b < 1$.*

Given these parameter restrictions, $\tau_b^{\exp}(x)$ and $\tau_b^{\log}(x)$ only include continuous and monotonically increasing functions, and, therefore, both deformations could also be used as a one-parameter continuum of KN functions. A useful relationship between concave deformations and quasilinear means is the following.

Lemma 13 (Deformation- and KN-Functions) *Let $\tau(x)$ be a concave function then $\tau(x) = c_1\phi^{-1}(x) + c_2$, $c_1 \neq 0 \Leftrightarrow \langle \tau(x) \rangle_\phi = \tau(\langle x \rangle_{\phi^{lin}})$ and $\tau(x) = c_1\phi(x) + c_2$, $c_1 \neq 0 \Leftrightarrow \langle x \rangle_\phi = \tau(\langle \tau^{-1}(x) \rangle_{\phi^{lin}})$.*

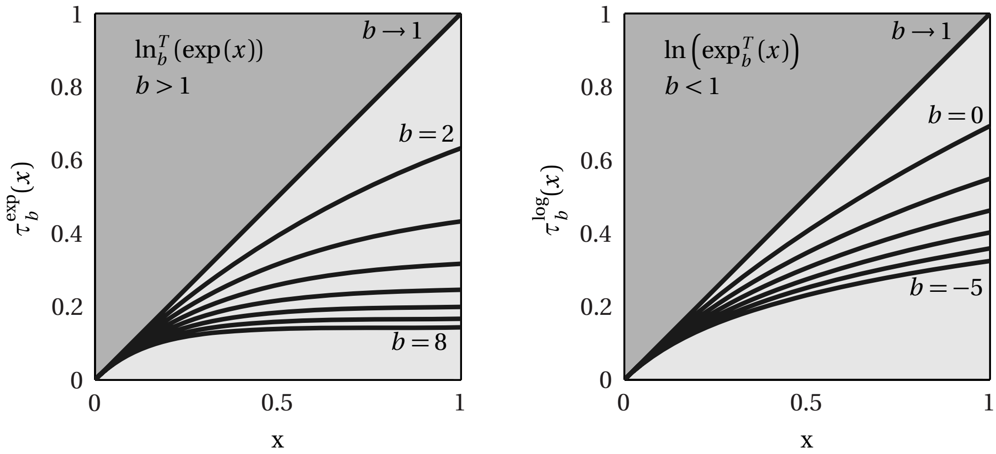

Figure 1. Deformations of x based on Tsallis' logarithm ln_b^T. On the left, the concave exponential deformation is shown ($b > 1$), on the right, the concave logarithmic deformation is shown ($b < 1$).

3 Quasilinear diversity

Let us consider the simplest of all KN functions, i.e. the linear one,

$$\phi^{\text{lin}}(x) = c_1 x + c_2, \ c_1 \neq 0 \tag{15}$$

then the quasilinear mean of rarity (6) generated by (15) is

$$\langle v_i \rangle_{\phi^{\text{lin}}} = V^S(\mathbf{p}). \tag{16}$$

We see that the Shannon diversity (7) can be expressed in terms of the quasilinear mean generated by an arbitrary linear KN function. This finding is a good starting point for generalization purposes. If we now allow any KN function to be used as mean generating function we get the *class of quasilinear diversities*

$$\langle v_i \rangle_{\phi} = \phi^{-1}\left(\sum_{i=1}^{n} p_i \phi(v_i)\right). \tag{17}$$

3.1 Order generalization of Shannon diversity

3.1.1 Derivation

Now, let us find the most general KN function that generates a diversity measure which i) is an element of (17) and ii) still has additivity property (2). Rényi (1961) proved that, besides linear KN functions (15), only linear transformations of one parameter exponentials are admissible, such that

$$\phi_a^{\exp}(x) = c_1 \exp\left((1-a)x\right) + c_2, \ c_1 \neq 0, \tag{18}$$

where a is the generalization parameter, called the *order*. It follows immediately from (6), (8) and (18) that [5]

$$\langle v_i \rangle_{\phi_a^{\exp}} = \frac{1}{1-a} \ln \left(\sum_{i=1}^{n} p_i \exp(v_i)^{1-a} \right) = \frac{1}{1-a} \ln \left(\sum_{i=1}^{n} p_i^a \right)$$
$$= \ln \left(\left(\sum_{i=1}^{n} p_i^a \right)^{\frac{1}{1-a}} \right) =: V_a^R(\mathbf{p}), \ a \neq 1, \ a \geq 0 \qquad (19)$$

This exponential mean of rarity v_i, usually called 'Rényi entropy' or 'Rényi diversity', recovers Shannon diversity for $a \to 1$ [6]. Comparing (18) with (13) we can easily see that $\phi_a^{\exp}(x) = \tau_a^{\exp}(x)$ for $c_1 = \frac{1}{1-b} = -c_2$ and, thus, $\langle v_i \rangle_{\phi_a^{\exp}} = \langle v_i \rangle_{\tau_a^{\exp}} = V_a^R(\mathbf{p})$ must hold. Due to Lemma 13 we may also write the Rényi diversity as logarithmically deformed function

$$V_a^R(\mathbf{p}) = \langle v_i \rangle_{\phi_a^{\exp}} = \tau_a^{\log} \left(\langle \tau_a^{\exp}(v_i) \rangle_{\phi^{\text{lin}}} \right) = \tau_a^{\log} \left(V_a^T(\mathbf{p}) \right), \qquad (20)$$

where $V_a^T(\mathbf{p})$ is another generalization of Shannon diversity, discussed in section 5.

3.1.2 Interpretation

The order a of Rényi diversity represents how much the differences between class sizes p_i are taken into account, when calculating the average rarity of the entire set of classes. Figure 2 shows that an order-0 diversity $v_0^R(\mathbf{p}) = \ln(n)$ does not weight at all. All n classes contribute equally to diversity, no matter how rare they actually are, and $v_0^R(\mathbf{p})$ is constant over all possible n-distributions $\mathbf{p}$. For positive orders smaller than 1 differences in the classes sizes p_i are more and more taken into account: The higher the size inequality between classes, the more weight is given to large classes compared to small ones. However, orders between zero and one imply that small classes are still given disproportionately high weights and large classes are given disproportionately low weights, compared to their actual sizes. Only order-1 diversity $v_1^R(\mathbf{p}) = \sum_{i=1}^n p_i v_i = V^S(\mathbf{p})$ generated by (15) weights the rarity v_i of the class i in exact proportion to the relative abundance p_i of that class. Finally, if the order a becomes larger than 1 then the large classes are weighted disproportionally high and small classes are weighted disproportionally low. This dis-

[5] Because negative orders usually do not make sense in the context of diversity measurement (cf. Baumgärtner 2002) we restrict our analysis to non-negative orders $a \geq 0$ even if some desirable properties may also be satisfied for $a < 0$.

[6] Note, that the subscript a will be replaced by other letters later on. To indicate that letters other than a still represent the order of the diversity we keep the superscript letter R (for Rényi).

proportionality increases with a, such that $\lim_{a\to\infty} V_a^R(\mathbf{p}) = -\ln(\max_i\{p_i\})$ only depends on the largest class available.

3.1.3 Additivity

Proposition 14 *For all $a \geq 0$ $V_a^R(\mathbf{p})$ is additive.*

3.1.4 Concavity

Proposition 15 *Let $\mathbf{p} \in \mathcal{C}$ and $n \geq 2$ then $V_a^R(\mathbf{p})$ is strictly concave for all $a \in (0, 1]$.*

Proposition 16 *Let $\mathbf{p} \in \mathcal{P}$ and $n = 2$ then $V_a^R(\mathbf{p})$ is strictly concave for all $a \in (0, 2)$.*

Proposition 17 *Let $\mathbf{p} \in \mathcal{P}$ and $n = 2$ then $V_a^R(\mathbf{p})$ is strictly concave for $a = 2$ iff $0 < p_1 < 1$.*

Proposition 18 *Let $n \geq 2$ and $a > 2$ then $V_a^R(\mathbf{p})$ is neither concave nor convex.*

4 Degree deformed diversity

Diversity of order a becomes non-concave for orders larger than two. That means, V_a^R is not an admissible one-parameter generalization of Shannon diversity when concavity is the primary requirement in a given context. Now, the aim is to find a generalization that maintains concavity for increasing generalization parameter.

4.1 Exponential deformation of Shannon diversity

4.1.1 Derivation

We know from proposition 11 that $\tau_b^{\exp}(x)$ is strictly concave for $b > 1$. Therefore, the *degree b-deformed* Shannon diversity

$$\begin{aligned} \tau_b^{\exp}(V^S(\mathbf{p})) &= \ln_b\left(\exp\left(V^S(\mathbf{p})\right)\right) \\ &= \frac{1}{1-b}\left(\exp\left((1-b)V^S(\mathbf{p})\right) - 1\right) =: V_b^G(\mathbf{p}) \end{aligned} \tag{21}$$

must preserve strict concavity for $b > 1$. V_b^G is indeed a generalization of Shannon diversity because $\lim_{b\to 1} V_b^G(\mathbf{p}) = V^S(\mathbf{p})$ immediately follows from (11). Lemma 10 indicates that a property "traded in" for concavity is additivity. To illustrate this more formally we replace the additivity axiom in theorem 4 by the non-additivity property (3) and obtain the functional equation $v_i(p_k p_l) = v_i(p_k) + v_i(p_l) + (1-b)v_i(p_k)v_i(p_l)$, $k, l \in \{1 \dots n\}, k \neq l$ which is equivalent to

$$1 + (1-b)v_i(p_k p_l) = (1 + (1-b)v_i(p_k))\,(1 + (1-b)v_i(p_l))\,. \tag{22}$$

Defining

$$f(p_k) := 1 + (1-b)v_i(p_k) \tag{23}$$

in (22) gives *Cauchy's power equation* $f(p_k p_l) = f(p_k)f(p_l)$, which is known to have the nonconstant and continuous solution [7]

$$f(p_k) = p_k^c,\ \ c \in \mathbb{R}\backslash\{0\}. \tag{24}$$

Now, let $c = b - 1$, $b \neq 1$ then follows the *b-deformed rarity* from (23) and (24):

$$v_{i,b} := \tau_b^{\exp}\left(v_i\right) = \ln_b^T\left(\frac{1}{p_i}\right). \tag{25}$$

Recalling Lemma 13, we can conclude that there is a direct link between $V_b^G(\mathbf{p})$ and the b-deformed rarity (25). Indeed, $V_b^G(\mathbf{p})$ can be expressed as a quasilinear mean of $v_{i,b}$ that is generated by any linear transformation of $\phi(x) = (\tau_b^{\exp}(x))^{-1} = \tau_b^{\log}(x)$:

$$V_b^G(\mathbf{p}) = \tau_b^{\exp}(\langle v_i\rangle_{\phi^{\mathrm{lin}}}) = \langle \tau_b^{\exp}\left(v_i\right)\rangle_{\phi_b^{\log}} \tag{26}$$

$$\text{where } \phi_b^{\log}(x) = c_1 \ln\left((1-b)\,x + 1\right) + c_2 = \tau_b^{\log}(x) \tag{27}$$

$$c_1 = \frac{1}{1-b},\ \ c_2 = 0.$$

4.1.2 Interpretation

Figure 2 illustrates some major differences between order and degree generalization of Shannon entropy. First, the maximum diversity $V_b^G(\bar{\mathbf{p}}) = \ln_b(n)$ now depends on the generalization parameter. Higher degrees imply smaller maxima. Second, and most important in the current context, concavity is preserved for increasing b. Only if the degree is decreased V_b^G may lose its concavity property, once having it. Third, the degree-0 diversity does not weight classes equally. Different distributions $\mathbf{p}$ imply very different values of $V_0^G(\mathbf{p})$. Such a property is difficult to interpret in the context of distribution based

[7] See Aczél et al. (2000) for an extensive review on functional equations and their solutions.

diversity measurement, which may be the reason why (21) has not been explicitly used as diversity measure so far. Nevertheless, we may call V_b^G the *Gaussian diversity*, simply to stay consistent with the nomenclature of other related measures [8]. Although Gaussian diversity may not have much practical relevance in most contexts of diversity measurement it is a quite meaningful example in generalized diversity theory.

4.1.3 Additivity

Proposition 19 *For all $b > 0$ $V_b^G(\mathbf{p})$ is non-additive of degree b.*

Proposition 20 *$V_b^G(\mathbf{p})$ is additive iff $b = 1$*

4.1.4 Concavity

Proposition 21 *Let $\mathbf{p} \in \mathcal{C}$ and $n \geq 2$ then $V_b^G(\mathbf{p})$ is strictly concave for all $b \geq 1$.*

Proposition 22 *Let $\mathbf{p} \in \mathcal{P}$ with $p_n = 1 - \sum_{i=1}^{n-1} p_i$ and $n \geq 2$ then $V_b^G(\mathbf{p})$ is strictly concave for all*

$$b \geq 1 - \frac{1 - p_n}{p_n \left(V^S(\mathbf{p}) + \ln(p_n)\right)^2}.$$

Example 23 (Approximation of concavity intervals) *Let $\mathbf{p} \in \mathcal{P}$ and $n = 2$ then $V_b^G(\mathbf{p})$ is strictly concave for all*

$$b \geq 1 - \frac{1}{(1-p_1)p_1 \ln\left(\frac{p_1}{1-p_1}\right)^2}. \tag{28}$$

Now let the second summand of inequality (28) be denoted $g(p_1)$. Since $\lim^+_{p_1 \to 0} g(p_1) = \infty$, $\lim^-_{p_1 \to 1} g(p_1) = \infty$ and $\lim_{p_1 \to 0.5} g(p_1) = \infty$, g is always positive and concavity is proved for $b \geq 1$. However, we can refine this concavity condition by finding the lower bound of g.

$$\begin{aligned} \frac{\partial g(p_1)}{\partial p_1} &= \frac{\ln\left(\frac{p_1}{1-p_1}\right)(1-2p_1) + 2}{p_1^2 + (1-p_1)^2 \ln\left(\frac{p_1}{1-p_1}\right)^3} \overset{!}{=} 0 \\ &\Rightarrow \ln\left(\frac{p_1}{1-p_1}\right)(2p_1 - 1) = 2 \end{aligned} \tag{29}$$

[8] For some negative degree V_b^G takes the typical "bell" shape of the Gaussian distribution.

Equation (29) is solved numerically by $p_1^* \approx 0.9168$ *and* $p_2^* = 1 - p_1^* \approx 0.0832$. *The second derivative of* g *is*

$$\frac{\partial^2 g(p_1)}{\partial p_1^2} = \frac{\overbrace{-6p_1\left(1-p_1\right)\ln\left(\frac{p_1}{1-p_1}\right)^2}^{>-3} + \overbrace{2\ln\left(\frac{p_1}{1-p_1}\right)\left(\ln\left(\frac{p_1}{1-p_1}\right)-6p_1+3\right)}^{>-2} + 6}{\underbrace{p_1^3\left(1-p_1\right)^3\ln\left(\frac{p_1}{1-p_1}\right)^4}_{\geq 0,\ \leq 0.02}} > 120 > 0$$

which means g *is minimal at* p_1^* *and* $1 - p_1^*$ *and has a lower bound of* $g(p_1^*) = g(1 - p_1^*) \approx 2.2767$. *Finally, from (28) follows the refined concavity condition* $b \gtrsim -1.2767$.

5 "Order = degree" generalization of Shannon diversity

We have seen that an order generalization of Shannon diversity preserves general additivity but removes general concavity, and deforming Shannon diversity exponentially by degree b preserves concavity but removes the general additivity property. This trade-off between valuable features cannot be fully avoided when generalizing Shannon diversity by quasilinear means or deformation. However, there are measures "in between" Rényi- and Gaussian diversity. For example, we may look for a diversity that is concave and non-additive, like Gaussian diversity, but still has the equivalent weighting property of order-0 diversity, $V_0^R(\mathbf{p}) = V_0^R(\mathbf{p}')$ for all $\mathbf{p} \neq \mathbf{p}'$ (cf. "Interpretation" in section 3.1).

5.1 Tsallis diversity

5.1.1 Derivation

The general idea is to make order-a diversity b-deformed, while order and degree depend in the same way on the same parameter b. If we apply the exponential deformation of degree b to the Rényi diversity and replace a with b we get with the help of Lemma 13

$$\tau_b^{\exp}\left(\langle v_i\rangle_{\phi_b^{\exp}}\right) = \langle v_{i,b}\rangle_{\phi^{\text{lin}}} = \sum_{i=1}^{n} p_i v_{i,b} = \sum_{i=1}^{n} p_i \tau_b(v_i)$$
$$= \frac{1}{1-b}\left(\sum_{i=1}^{n} p_i^b - 1\right) =: V_b^T(\mathbf{p}),\ b \geq 0,\ b \neq 1. \tag{30}$$

Obviously, $V_b^T(\mathbf{p})$ can either be seen as a b-deformed Rényi diversity or equivalently as a Gaussian diversity in which the logarithmic mean generation (27) is replaced with the linear one (15). Therefore, we may call $V_b^T(\mathbf{p})$ the *diversity of order and degree* b, or, due to its most famous creator Tsallis (1988), *Tsallis diversity.*

5.1.2 Interpretation

As figure 2 illustrates, Tsallis diversity $V_b^T(\mathbf{p})$ unifies two desirable properties of Rényi and Gaussian diversity. The property $V_0^T(\mathbf{p}) = v_0^T(\mathbf{p}')$ for all $\mathbf{p} \neq \mathbf{p}'$ is also satisfied by 'pure' order generalization $V_a^R(\mathbf{p})$ and the preservation of concavity for increasing parameter values is also satisfied by 'pure' degree deformation $V_b^G(\mathbf{p})$. Note further that $V_b^T(\bar{\mathbf{p}}) = V_b^G(\bar{\mathbf{p}}) = \ln_b(n)$ is another common feature of Tsallis and Gaussian diversity.

Due to its unification of desirable properties, Tsallis diversity is, in contrast to Gaussian diversity, quite appealing for practical diversity measurement. Thus, it is not very suprising that scientists once in a while "re-invent" V_b^T as appropriate diversity measure (e.g. Keylock 2005). More surprising is the observation that the equivalence between Tsallis diversity and Patil and Taillie's (1982) quite popular "diversity index of degree β" seems to be largely unknown [9]. In fact, using V_b^T as generalized diversity measure is at least two decades old.

5.1.3 Additivity

Proposition 24 *For all $b > 0$ $V_b^T(\mathbf{p})$ is non-additive of degree b.*

Proposition 25 *$V_b^T(\mathbf{p})$ is additive iff $b = 1$*

5.1.4 Concavity

Proposition 26 *Let $n \geq 2$, $p \in \mathcal{C}$ or $p \in \mathcal{P}$ then $V_b^T(\mathbf{p})$ is strictly concave for all $b > 0$.*

[9] It can be obtained by simple paramter transformation $b \leftrightarrow \beta - 1$ which is, graphically seen, a horizontal left-shift by 1 of the curves in figure 2.

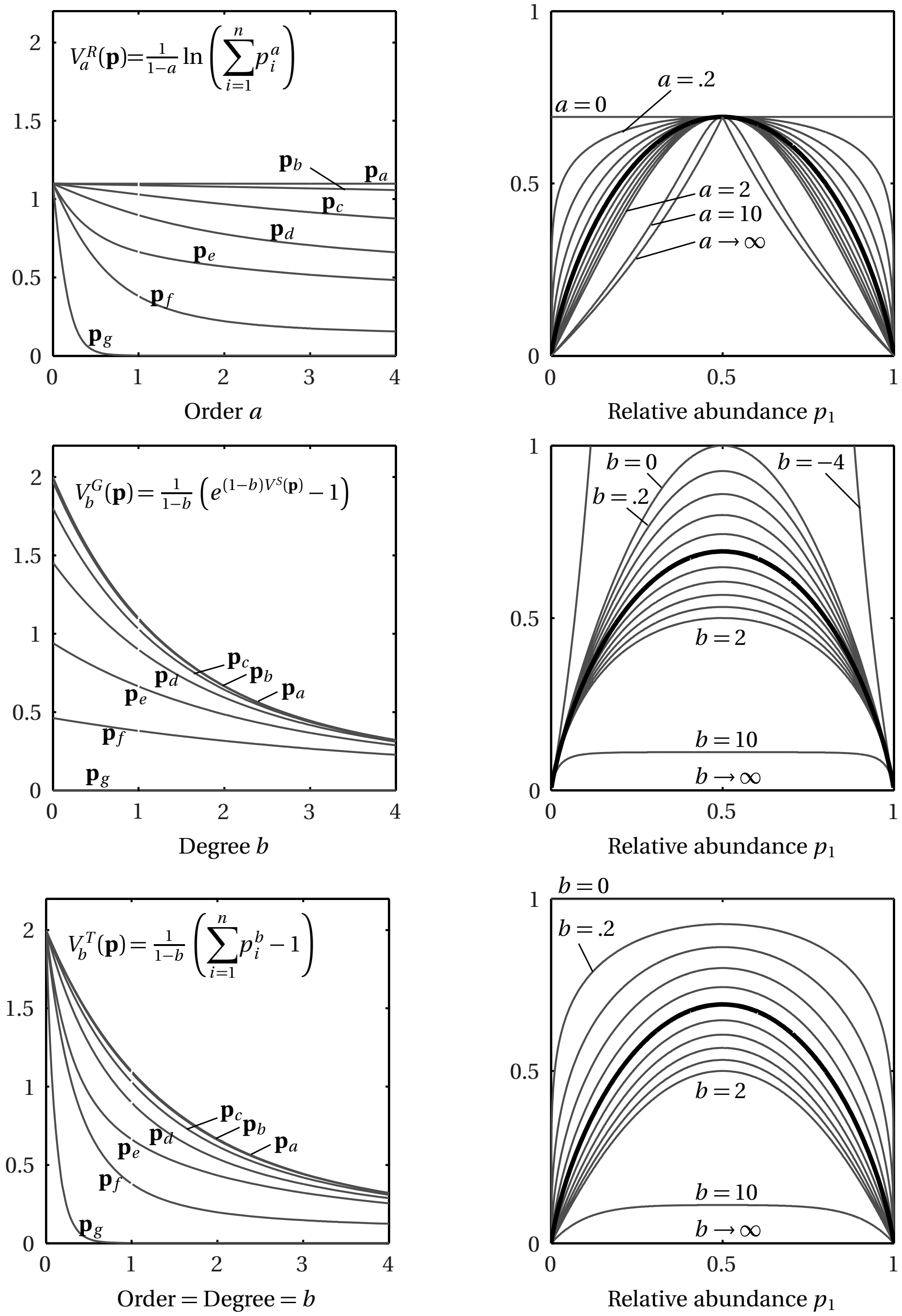

Figure 2. Graphical representation of different generalizations of Shannon diversity. The two columns show the generalizations depending on the generalization parameter and p_1, ($n = 2$) respectively. Distributions $\mathbf{p}_a$ to $\mathbf{p}_g$ are taken from table 2, p. 21. In the right column the Shannon diversity is marked bold.

Diversity	Generalization	Representation	Argument	τ-defm.	ϕ-weight.
Shannon	-	$\langle v_i \rangle_{\phi^{\text{lin}}}$	rarity	-	lin.
Rényi	order	$\langle v_i \rangle_{\phi_a^{\exp}}$	rarity	-	exp.
		$\tau_a^{\log}\left(V_a^T\right)$	diversity	log.	-
Gauss	degree	$\left\langle \tau_b^{\exp}\left(v_i\right) \right\rangle_{\phi_b^{\log}}$	defm. rarity	-	log.
		$\tau_b^{\exp}\left(V^S\right)$	diversity	exp.	-
Tsallis	order = degree	$\left\langle \tau_b^{\exp}\left(v_i\right) \right\rangle_{\phi^{\text{lin}}}$	defm. rarity	exp.	lin.
		$\tau_b^{\exp}\left(V_b^R\right)$	diversity	exp.	-

Table 1
Three kinds of generalization of the Shannon diversity.

6 "Order $\neq$ degree" generalization of Shannon diversity

6.1 Sharma-Mittal diversity

6.1.1 Derivation

Table 1 summarizes the generalizations of Shannon diversity discussed so far, including possible transformations between them. The rows show how to derive the generalization, whereas the last three columns can give an idea of how to find a unifying notation of all three generalized diversities V_a^R, V_b^G and V_b^T. In the last column of table 1 we have linear, logarithmic and expontential KN functions. A first approach could be to introduce a two parameter KN function that includes all three weighting schemes as special cases. Obviously, the two parameter deformation $\tau_{a,b}(x) := \ln_a^T(\exp_b^T(x))$ becomes $\tau_b^{\log}(x)$ for $a \to 1$, $\tau_b^{\exp}(x)$ for $b \to 1$ and linear for $a = b$. Thus, the two parameter KN function

$$
\begin{aligned}
\phi_{a,b}(x) &:= c_1 \ln_a^T(\exp_b^T(x)) + c_2 \\
\phi_{a,b}^{-1}(x) &= c_1 \ln_b^T(\exp_a^T(x)) + c_2
\end{aligned}
$$

covers all deformations (column 5) and KN functions (column 6) of table 1:

$$\phi_{a,a}(x) = \phi_{b,b}(x) = c_1 x + c_2 = \phi^{\text{lin}} \text{ (Tsallis and Shannon)}$$
$$\lim_{a\to 1} \phi_{a,b}(x) =: \phi_{1,b}(x) = c_1 \ln(\exp_b^T(x)) + c_2 = \phi_b^{\log} \text{ (Gauss)}$$
$$\lim_{b\to 1} \phi_{a,b}(x) =: \phi_{a,1}(x) = c_1 \ln_a^T(\exp(x)) + c_2 \ = \phi_a^{\exp} \text{ (Rényi)}.$$

The arguments of the quasilinear mean representations (Column 4) are generalized by the deformed rarity (25) and, therefore, we calculate the quasilinear mean of $v_{i,b}$ with respect to $\phi_{a,b}$. With the help of Lemma 13 we quickly get

$$\langle v_{i,b}\rangle_{\phi_{a,b}} = \phi_{a,b}^{-1}\left(\sum_{i=1}^{n} p_i \phi_{a,b}\left(\ln_b^T\left(\frac{1}{p_i}\right)\right)\right) = \ln_b^T\left(\exp_a^T\left(V_a^T(\mathbf{p})\right)\right) \tag{31}$$

$$= \frac{1}{1-b}\left(\left(\sum_{i=1}^{n} p_i^a\right)^{\frac{1-b}{1-a}} - 1\right) =: V_{a,b}^{SM}(\mathbf{p}). \tag{32}$$

This two parameter generalization of Shannon diversity was derived first (in another context and a slightly modified way) by Sharma and Mittal (1975, 1977). We call it the *diversity of order a and degree b* [10], or simply *Sharma-Mittal diversity* because of the following properties:

$$\lim_{a\to 1} V_{a,b}^{SM}(\mathbf{p}) = \frac{1}{1-b}\left(e^{(1-b)V^S(\mathbf{p})} - 1\right) = V_b^G(\mathbf{p}) =: V_{1,b}^{SM}(\mathbf{p}) \tag{33}$$

$$\lim_{b\to 1} V_{a,b}^{SM}(\mathbf{p}) = \frac{1}{1-a}\ln\left(\sum_{i=1}^{n} p_i^a\right) = V_a^R(\mathbf{p}) =: V_{a,1}^{SM}(\mathbf{p}) \tag{34}$$

$$\lim_{a\to 1}\lim_{b\to 1} V_{a,b}^{SM}(\mathbf{p}) = \sum_{i=1}^{n} p_i \ln\left(\frac{1}{p_i}\right) = V^S(\mathbf{p}) =: V_{1,1}^{SM}(\mathbf{p}) \tag{35}$$

$$V_{a=b}^{SM}(\mathbf{p}) = \frac{1}{1-b}\left(\sum_{i=1}^{n} p_i^b - 1\right) = V_b^T(\mathbf{p}) =: V_{b,b}^{SM}(\mathbf{p}). \tag{36}$$

6.1.2 Interpretation

Combining (20) with (31) reveals that the Sharma-Mittal diversity essentially is a b-deformed diversity of order a. By simply replacing $\ln(x)$ with $\ln_b^T(x)$ in (19) we can verify indeed that

$$V_{a,b}^{SM}(\mathbf{p}) = \tau_b^{\exp}(V_a^R). \tag{37}$$

Because the Sharma-Mittal diversity is a generalization of V_a^R, V_b^G and V_b^T, it unifies the typical characteristics identified so far in a single framework. Most important are (cf. figure 3)

[10] In information theory $H_{a,b}^{SM}$ is usually called *information of order a and degree b* (Taneja, 1989) or *information of order a und rank b* (Aczél, 1984).

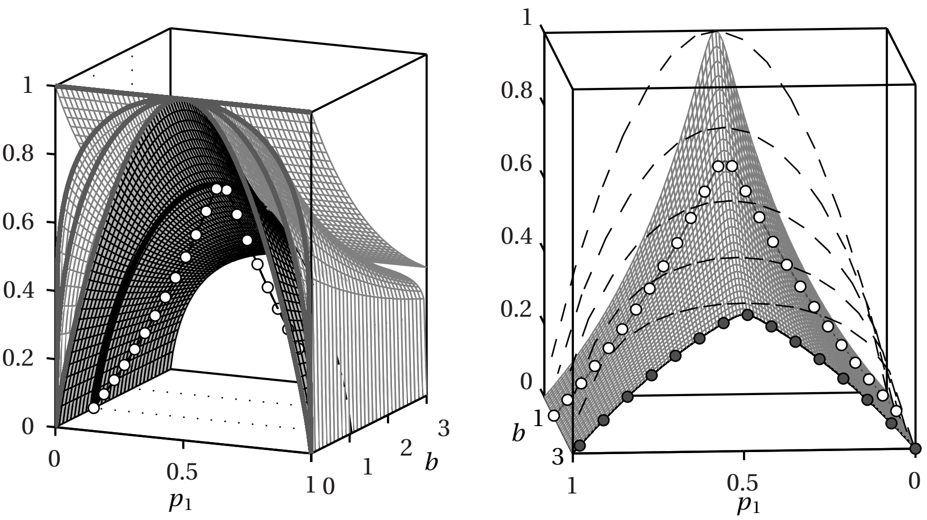

Figure 3. Diversity of order a and degree b given $p = (p_1, p_2)$ where $p_2 = 1 - p_1$. Here, order $a \in \{0, 0.2, 0.4, 1\}$ is used as discrete parameter which causes different layers of the left plot. The darkened grid repesents all $a = 1$ diversities V_b^G and the Shannon diversity, again, is marked black-bold. To provide a better comparability with the right plot, the non-concave diversity $V_{20,1}^{SM}$ is marked as a white dotted line. The right plot shows diversity $V_{20,b}^{SM}$ and illustrates the transformation of a strictly concave diversity (dark dotted) to a non-concave and non-convex diversity (white dotted) depending on b. The Tsallis case $b = a$, i.e. the all-concave V_b^T, is marked dashed.

- $V_{0,b}^{SM}(\mathbf{p}) = V_{0,b}^{SM}(\mathbf{p}')$ for all $\mathbf{p} \neq \mathbf{p}'$ and all $b \geq 0$ (equivalent weighting property of order-0 diversities).
- Preservation of concavity of order-1 diversities for increasing degree of deformation b.
- Monotonically decreasing in order a and degree b.
- Non-negative for all orders a and degrees b.
- Maximum of $V_{0,b}^{SM}(\bar{\mathbf{p}}) = \log_b^T(n)$ (order-0 maximal)

Moreover, it is quite interesting to see that the degree 0-deformation of Sharma-Mittal diversity is equivalent to

$$V_{a,0}^{SM}(\mathbf{p}) = N_a - 1,$$

where N_a is the very classic 'effective number' N_a (of classes) introduced by Hill (1973) and recently named 'true' diversity by Jost (2006). In economics N_a is widely used to measure (inverse) concentration of an industry (Hannah and Kay, 1977), in political sciences this number has established as standard to measure the 'effective number of parties' (Laakso and Taagepera, 1979). However, the major benefits of N_a, i.e. a straightforward interpretation of its unit and a nice "doubling" property (Jost, 2006), again, come at some costs. It can be easily shown that N_a is neither additive, nor generally concave (cf. figure 4). In other words, for the sake of a straightforward interpretation (of the unit of measurement) the 'effective number'-transformation of entropic measures is very reasonable, for the sake of two common diversity properties it is not [11].

As usual, concavity and additivity are discussed separately.

6.1.3 Additivity

Proposition 27 *For all $b > 0$ $V_{a,b}^{SM}(\mathbf{p})$ is non-additive of degree b.*

Proposition 28 *$V_{a,b}^{SM}(\mathbf{p})$ is additive iff $b = 1$*

6.1.4 Concavity

Proposition 29 *Let $\mathbf{p}_i \in \mathcal{C}$ then $V_{a,b}^{SM}(\mathbf{p})$ is strictly concave for all*

$$b > 1 - \frac{1-a}{a}.$$

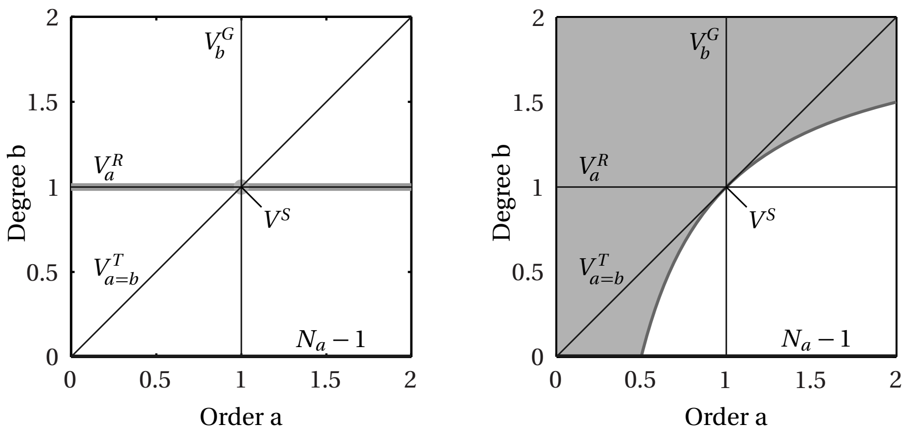

Figure 4. Additivity-diagram (left) and concavity-diagram (right) of diversity measures in the Sharma-Mittal parameter space. The thin black lines represent generalized one-parameter diversities as a function of their parameter and the greyed areas represent parameter combinations resulting in addititive and concave measures respectively.

[11] Therefore Jost's (2006) naming can obviously be quite misleading and should be used with care (cf. Hoffmann and Hoffmann 2006).

7 Conclusion

As old as the Sharma-Mittal formalism is, as unknown it seems to be in applied diversity theory. This is not very reasonable. Regarding the fact that Rényi's (1961) order a entropy is indeed a very popular diversity measure in ecology, economics and other sciences, the b-deformed Rényi diversity $V_{a,b}^{SM}$ represents nothing but a logical step towards a more flexible and property-driven modelling of distribution based diversity, and, therefore, desires more attention apart from information theory and statistical mechanics. In other words, if some scientists find the additive diversity class V_a^R admissible (e.g. Pielou 1975) but others prefer the concave diversity class V_b^T (e.g. Keylock 2005) both point of views can be captured by the same quantitative model $V_{a,b}^{SM}$. By setting the parameters of Sharma-Mittal diversity accordingly to individual judgements on the given application context we can establish a profound methodological reasoning in diversity measurement. Moreover it becomes much easier to analyze the interplay between important diversity properties, as it was shown for additivity and concavity.

8 Outlook

The analysis of this paper is limited in two ways. First, only two properties, additivity and concavity, are analyzed and compared. Many more properties may be important for the measurement of distribution based diversity [12]. However, both properties were delibarately chosen, since they belong indeed to the most "typical" ones and they allow to illustrate a typical trade-off between properties that is implied by common generalization techniques. Moreover, analyzing a more exhaustive list of properties in the same manner would quickly go beyond the scope of this paper. The second limitation of our analysis refers to the Tsallis logarithm (9). Other deformed logarithms exist and other one- and two-parameter generalizations of Shannon diversity can be derived from them. Abe (1997), for example, presents a b-deformed logarithm having the $b \leftrightarrow 1/b$ invariance property. His resulting entropy recovers Shannon entropy for $b \to 1$ like most other generalizations. More recently, Kaniadakis and Scarfone (2002) introduced the κ-deformed-logarithm

$$\begin{aligned} \ln_\kappa(x) &= \frac{x^\kappa - x^{-\kappa}}{2\kappa} \\ \lim_{\kappa \to 0} \ln_\kappa(x) &= \ln(x) \end{aligned}$$

[12] Several properties of Shannon entropy and some of its derivatives can be found in Aczél and Daróczy (1975) and Taneja (1989).

which has the property $\ln_\kappa(1/x) = -\ln_\kappa(x)$ known from standard algebra. Kaniadakis et al. (2005) further extend the κ-deformed-logarithm to the $\kappa - r$-deformed logarithm $\ln_{\kappa,r}(x) = x^r \ln_\kappa(x)$. Their two parameter Shannon generalization $H_{\kappa,r}(\mathbf{p}) = -\sum p_i \ln_{\kappa,r}(p_i)$ is equivalent to an entropy measure first introduced in physics by Borges and Roditi (1998) and also known as *entropy of type (a,b)* (Sharma and Taneja, 1975) or *entropy of degree (a,b)* (Aczél, 1984) in information theory [13]. Tsallis entropy and Abe entropy are prominent special cases of these generalizations.

Information theory and physics have definitely proved to be rich sources of knowledge when the aim is to find generalized measures of distribution based diversity. However, the analogy of this very special kind of diversity and entropy measures is not perfect, and the interdisciplinary transfer of measurement concepts should not happen blindly. For example, the class $H_{\kappa,r}$ is not additive. This drawback may already indicate that it is not an admissible class of *diversity* measures, *although* $H_{\kappa,r}$ may be truely admissible to measure entropy, information or other phenomenological quantities. To the best of our knowledge the Tsallis-logarithms is most appropriate in order to develop a unifying notation of frequently used diversiy measures, that can also serve as a “property comparison framework”. But at the end, problems like whether or not an unknown concept can serve as a diversity measure, always narrow down to two central questions: First, “What properties should a diversity measure have?”, or, more generally speaking, “What is diversity (in the given context)?”, and second, “What properties does the concept in question provide?”. The better the matchup of given answers, the higher the legitimation to use that concept. Unfortunately these questions are often all but easy to answer and a lot of future research on the measurement of diversity still needs to be carried out to fully understand this important phenomenon and its multidisciplinary connections.

[13] The Borges/Roditi formalism is obtained by simple paramter transformation $\kappa \leftrightarrow (\beta - \alpha)/2$ and $r = (\alpha + \beta)/2 - 1$. In information theory the entropy of type (a, b) is usually written in a normalized form (cf. Kapur 1994, ch. 18).

Appendix

8.1 Some hypothetical distributions

$\mathbf{p}_a$	0.33	0.33	0.33
$\mathbf{p}_b$	0.3	0.3	0.4
$\mathbf{p}_c$	0.2	0.3	0.5
$\mathbf{p}_d$	0.1	0.3	0.6
$\mathbf{p}_e$	0.01	0.3	0.69
$\mathbf{p}_f$	0.01	0.1	0.89
$\mathbf{p}_g$	0.00005	0.00005	0.9999

Table 2
Seven different abundance distributions of $n = 3$ given classes.

Proofs

8.1.1 Section 2

Proof of Theorem 4, p. 4. Aczél and Daróczy (1975), p. 3 ■

Proof of Proposition 5, p. 5. Aczél and Daróczy (1975), p. 31 ■

Proof of Proposition 6, p. 5. Let $p \in \mathcal{C}$, then the symmetric $n \times n$ Hessian $\mathbf{H}^{V^S}$ is:

$$\mathbf{H}^{V^S} = \begin{pmatrix} -\frac{1}{p_1} & 0 & \dots & 0 \\ & -\frac{1}{p_2} & \dots & 0 \\ & & \ddots & \vdots \\ & & & -\frac{1}{p_n} \end{pmatrix}$$

Concavity criterion (5) implies $\mathbf{p}(-\mathbf{H}^{V^S})\mathbf{p}' = \sum_{i=1}^{n} p_i = 1 > 0$. ■

Proof of Theorem 8, p. 5. Hardy et al. (1934), p. 66-67 ■

Proof of Lemma 10, p. 6. Let $\mathbf{p}_i$ and $\mathbf{p}_j$ be two independent distributions and suppose $\tau_b^{\exp}(V(\cdot))$ is non-additive of degree b then

$$
\begin{aligned}
\tau_b^{\exp}(V(\mathbf{p}_{ij})) &= \tau_b^{\exp}(V(\mathbf{p}_i)) + \tau_b^{\exp}(V(\mathbf{p}_j)) \\
&\quad + (1-b)\,\tau_b^{\exp}(V(\mathbf{p}_i))\tau_b^{\exp}(V(\mathbf{p}_j)) \\
\exp\left((1-b)\,V(\mathbf{p}_{ij})\right) + 1 &= \exp\left((1-b)\,V(\mathbf{p}_i)\right)\exp\left((1-b)\,V(\mathbf{p}_j)\right) \\
&\quad + \left(\exp\left((1-b)\,V(\mathbf{p}_i)\right) - 1\right)\left(\exp\left((1-b)\,V(\mathbf{p}_j)\right) - 1\right) \\
\exp\left((1-b)\,V(\mathbf{p}_{ij})\right) &= \exp\left((1-b)\,V(\mathbf{p}_i)\right)\exp\left((1-b)\,V(\mathbf{p}_j)\right) \\
V(\mathbf{p}_{ij}) &= V(\mathbf{p}_i) + V(\mathbf{p}_j),
\end{aligned}
$$

which is true iff V is additive. ■

Proof of Proposition 11, p. 6. The second derivative is $\frac{\partial^2 \tau_b^{\exp}(x)}{\partial x^2} = (1-b)\exp((1-b)x)$ which is negative for all $b > 1$, because $\exp(\cdot)$ is strictly positive. ■

Proof of Proposition 12, p. 6. The second derivative is $\frac{\partial^2 \tau_b^{\log}(x)}{\partial x^2} = -(1-b)\cdot((1-b)x+1)^{-2}$ which is negative for all $b < 1$, because $-(1-b)$ is negative and $((1-b)x+1)^{-2}$ is positive for $b < 1$. ■

Proof of Lemma 13, p. 6. Both cases follow from definition 8 and theorem 8:

(1) If $\tau(x) = \phi^{-1}(x)$ then $\langle \tau(x) \rangle_\phi = \phi^{-1}\left(\sum_i p_i \phi\left(\tau(x)\right)\right) = \tau\left(\sum_i p_i x\right) = \tau\left(\langle x \rangle_{\phi^{\text{lin}}}\right)$
(2) If $\tau(x) = \phi(x)$ then $\langle x \rangle_\phi = \phi^{-1}\left(\sum_i p_i \phi\left(x\right)\right) = \tau^{-1}\left(\sum_i p_i \tau\left(x\right)\right) = \tau^{-1}(\langle \tau(x) \rangle_{\phi^{\text{lin}}})$

■

8.1.2 Section 3

Proof of Proposition 14, p. 9. Let $\mathbf{p}_i$ and $\mathbf{p}_j$ be two independent distributions then

$$
\begin{aligned}
V_a^R(\mathbf{p}_{ij}) &= \frac{1}{1-a}\ln\left(\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^a\right) \\
&= \frac{1}{1-a}\ln\left(\sum_{i=1}^{n}\sum_{j=1}^{m} p_i^a p_j^a\right) = \frac{1}{1-a}\ln\left(\sum_{i=1}^{n} p_i^a \sum_{j=1}^{m} p_j^a\right) \\
&= \frac{1}{1-a}\ln\left(\sum_{i=1}^{n} p_i^a\right) + \frac{1}{1-a}\ln\left(\sum_{j=1}^{m} p_j^a\right) \\
&= V_a^R(\mathbf{p}_i) + V_a^R(\mathbf{p}_j)
\end{aligned}
$$

Rényi diversity is additive for all $a \in \mathbb{R}$ which proves proposition 14. ■

Proof of Proposition 15, p. 9. For $a \to 1$ we have V^S which is known to be strictly concave with respect to $\mathbf{p}$. For all $a \in (0,1)$ we have $\mathbf{p}(-\mathbf{H}^{V_a^R})\mathbf{p}' = \frac{a}{1-a} > 0$. ■

Proof of Proposition 16 and 17, p. 9. Proposition 16 was proved by Ben-Bassat and Raviv (1978). Here we give a slightly corrected version of their proof. Let $\mathbf{p} = (p_1, 1-p_1)$ then

$$\frac{\partial^2 V_a^R(\mathbf{p})}{\partial p_1^2} = \frac{a}{1-a} \frac{\left(a p_1^{a-2}(1-p_1)^{a-2} - \left(p_1^{a-2} + (1-p_1)^{a-2}\right)\left(p_1^a + (1-p_1)^a\right)\right)}{\left(p_1^a + (1-p_1)\right)^2}$$

Since $a/(1-a) < 0$ for all $a > 1$ and $\left(p_1^a + (1-p_1)\right)^2 > 0$ the condition for strict concavity can be written as

$$a p_1^{a-2}(1-p_1)^{a-2} - \left(p_1^{a-2} + (1-p_1)^{a-2}\right)\left(p_1^a + (1-p_1)^a\right) > 0.$$

Dividing by $p_1^{a-2} + (1-p_1)^{a-2} > 0$ gives

$$a - \left(p_1^{2-a} + (1-p_1)^{2-a}\right)\left(p_1^a + (1-p_1)^a\right) > 0, \tag{38}$$

where $p_1^a + (1-p_1)^a < 1$ for all $a > 1$ such that

$$a - \left(p_1^{2-a} + (1-p_1)^{2-a}\right) > 0. \tag{39}$$

Now $p_1^{2-a} + (1-p_1)^{2-a}$ is maximal at $p_1 = 1 - p_1 = \frac{1}{2}$ and therefore the concavity condition (39) can be written as

$$a - \left(\frac{1}{2}\right)^{1-a} > 0,$$

which is true for $1 < a < 2$. For $a = 2$ the strict concavity condition (38) holds iff

$$p_1 > p_1^2$$

which is true for all $0 < p_1 < 1$. ■

Proof of Proposition 18, p. 9. Ben-Bassat and Raviv (1978). ■

8.1.3 Section 4

Proof of Proposition 19, p. 11. Follows immediately from (21) and lemma 10. ■

Proof of Proposition 20, p. 11. Follows immediately from proposition 19). ■

Proof of Proposition 21, p. 11. Let $p \in \mathcal{C}$, $r = 1 \dots n$ be the line index (vertical direction) and $s = 1 \dots n$ the row index (horizontal direction) of the symmetric $n \times n$ Hessian $\mathbf{H}^{V_b^G}$. Further define $y_r(\mathbf{p}) := \ln(p_r) + 1$, $z(\mathbf{p}) := e^{(1-b)V^S(\mathbf{p})}$ then

$$\mathbf{H}^{V_b^G} = \begin{pmatrix} \frac{z(\mathbf{p})}{p_1} + (b-1)\, y_1(\mathbf{p})^2 z(\mathbf{p}) & (b-1)\, y_1(\mathbf{p})\, y_2(\mathbf{p})\, z(\mathbf{p}) & \dots & (b-1)\, y_1(\mathbf{p})\, y_n(\mathbf{p})\, z(\mathbf{p}) \\ & \frac{z(\mathbf{p})}{p_2} + (b-1)\, y_2(\mathbf{p})^2 z(\mathbf{p}) & \dots & (b-1)\, y_2(\mathbf{p})\, y_n(\mathbf{p})\, z(\mathbf{p}) \\ & & \ddots & \vdots \\ & & & \frac{z(\mathbf{p})}{p_n} + (b-1)\, y_n(\mathbf{p})^2 z(\mathbf{p}) \end{pmatrix}.$$

Concavity criterion (5) implies

$$\begin{aligned} \mathbf{p}(-\mathbf{H}^{V_b^G})\mathbf{p}' &= \underbrace{z(\mathbf{p})}_{>0} \left[\underbrace{\sum_{r=1}^{n} p_r}_{1} + (b-1) \cdot \underbrace{\sum_{r=1}^{n} \sum_{s=1}^{n} p_r y_r(\mathbf{p})\, p_s y_s(\mathbf{p})}_{(V^S(\mathbf{p})-1)^2} \right] \overset{!}{>} 0 \\ \Rightarrow b &\overset{!}{>} 1 - \frac{1}{(V^S(\mathbf{p}) - 1)^2} =: g(V^S(\mathbf{p})) \end{aligned} \tag{40}$$

From the well-known inequality $0 \leq V^S(\mathbf{p}) \leq \ln(n)$, $n \geq 2$ and $\lim_{n\to\infty} g(\ln(n)) = \lim_{n\to 0} g(\ln(n)) = 1$ follows $0 \leq g(V^S(\mathbf{p})) \leq g(\ln(n)) < 1$. Then, (40) is obviously true iff $b \geq 1$. $\blacksquare$

Proof of Proposition 22, p. 11. Let $p \in \mathcal{P}$, where $p_n = 1 - \sum_{i=1}^{n-1} p_i$, $r = 1 \dots n-1$ be the line index (vertical direction) and $s = 1 \dots n-1$ the row index (horizontal direction) of the symmetric $(n-1) \times (n-1)$ Hessian $\mathbf{H}^{V_b^G}$.

Further define $y_r(\mathbf{p}) := \ln\left(\frac{p_r}{p_n}\right)$, $z(\mathbf{p}) := e^{(1-b)V^S(\mathbf{p})}$ then

$$\mathbf{H}^{V_b^G} =$$
$$\begin{pmatrix} \left(\frac{1}{p_1}+\frac{1}{p_n}\right) z(\mathbf{p}) + (b-1)\, y_1(\mathbf{p})^2 z(\mathbf{p}) & (b-1)\, y_1(\mathbf{p})\, y_2(\mathbf{p})\, z(\mathbf{p}) & \dots & (b-1)\, y_1(\mathbf{p})\, y_{n-1}(\mathbf{p})\, z(\mathbf{p}) \\ & \left(\frac{1}{p_2}+\frac{1}{p_n}\right) z(\mathbf{p}) + (b-1)\, y_2(\mathbf{p})^2 z(\mathbf{p}) & \dots & (b-1)\, y_2(\mathbf{p})\, y_{n-1}(\mathbf{p})\, z(\mathbf{p}) \\ & & \ddots & \vdots \\ & & & \left(\frac{1}{p_2}+\frac{1}{p_n}\right) z(\mathbf{p}) + (b-1)\, y_{n-1}(\mathbf{p})^2 z(\mathbf{p}) \end{pmatrix}$$

Concavity criterion (5) implies

$$\mathbf{p}(-\mathbf{H}^{V_b^G})\mathbf{p}' =$$
$$\underbrace{z(\mathbf{p})}_{>0}\left[\underbrace{\sum_{r=1}^{n-1} p_r\left(1+\frac{\sum_{r=1}^{n-1} p_r}{p_n}\right)}_{\frac{1-p_n}{p_n}} + (b-1)\cdot\underbrace{\left(\ln(p_n)\sum_{r=1}^{n-1} p_r - \sum_{r=1}^{n-1} p_r \ln(p_r)\right)^2}_{(V^S(\mathbf{p})+\ln(p_n))^2}\right] \overset{!}{>} 0$$
$$\Rightarrow b > 1 - \frac{1-p_n}{p_n\left(V^S(\mathbf{p})+\ln(p_n)\right)^2}$$

■

8.1.4 Section 5

Proof of Proposition 24, p. 13. Because $V_b^T(\mathbf{p}) = \tau_b^{\exp}\left(V_b^R(\mathbf{p})\right)$ (cf. (30), p. 13) non-aditivity of degree b follows from lemma 10, p. 6 and proposition 14, p. 14. ■

Proof of Proposition 25, p. 13. Follows immediately from proposition 24. ■

Proof of Proposition 26, p. 13. If $p \in \mathcal{C}$ we have $\mathbf{p}(-\mathbf{H}^{v_a^T})\mathbf{p}' = b\sum_{i=1}^{n} p_i^b > 0$ for all $b > 0$ and if $\mathbf{p} \in \mathcal{P}$, $p_n = 1 - \sum_{i=1}^{n-1} p_i$ we have $\mathbf{p}(-\mathbf{H}^{v_a^T})\mathbf{p}' = b\left(\sum_{i=1}^{n-1} p_i^b + (1-p_n)^2 p_n^{b-2}\right) > 0$ for all $b > 0$. ■

8.1.5 Section 6

Proof of Proposition 27, p. 18. Because $V_{a,b}^{SM}(\mathbf{p}) = \tau_b^{\exp}\left(V_a^R(\mathbf{p})\right)$ (cf. (37), p. 16) non-aditivity of degree b follows from lemma 10, p. 6 and proposition 14, p. 14. ■

Proof of Proposition 28, p. 18. Follows immediately from proposition 27. ■

Proof of Proposition 29, p. 18. Let $p \in \mathcal{C}$ and $r = 1 \ldots n-1$ be the line index (vertical direction) and $s = 1 \ldots n-1$ the row index (horizontal direction) of the symmetric $n \times n$ Hessian $\mathbf{H}^{V_{a,b}^{SM}}$. Further define $y(\mathbf{p}) := \sum_{r=1}^{n} p_r^a$, $z(\mathbf{p}) := -\frac{a}{(1-a)^2} y(\mathbf{p})^{\frac{1-b}{1-a}-2}$, $\mathfrak{a} = ab - a^2$ and $\mathfrak{b} = (a-1)^2$ then

$$\mathbf{H}^{V_{a,b}^{SM}} =$$

$$\begin{pmatrix} z(\mathbf{p})\, p_1^{a-2}\left(\mathfrak{a} p_1^a + \mathfrak{b} y(\mathbf{p})\right) & \mathfrak{a} z(\mathbf{p})\, p_1^{a-1} p_2^{a-1} & \ldots & \mathfrak{a} z(\mathbf{p})\, p_1^{a-1} p_n^{a-1} \\ & z(\mathbf{p})\, p_2^{a-2}\left(\mathfrak{a} p_2^a + \mathfrak{b} y(\mathbf{p})\right) & \ldots & \mathfrak{a} z(\mathbf{p})\, p_2^{a-1} p_n^{a-1} \\ & & \ddots & \vdots \\ & & & z(\mathbf{p})\, p_n^{a-2}\left(\mathfrak{a} p_n^a + \mathfrak{b} y(\mathbf{p})\right) \end{pmatrix}$$

Concavity criterion (5) implies

$$\mathbf{p}(-\mathbf{H}^{V_{a,b}^{SM}})\mathbf{p}' = \frac{a(1-2a+ab)}{\underbrace{(a-1)^2}_{>0}} \underbrace{\left(\sum_{i=1}^{n} p_i^a\right)^{\frac{1-b}{1-a}}}_{>0} \overset{!}{>} 0$$

$$\Rightarrow a(1-2a+ab) > 0$$

$$\Rightarrow b > \frac{2a-1}{a} = 1 - \frac{1-a}{a}$$

■

Acknowledgements

I thank Joachim Weimann, Bodo Vogt, Thomas Riechmann and Stefan Baumgärtner for helpful comments. Financial support from *Graduiertenförderung der Otto-von-Guericke-Universität Magdeburg* is gratefully acknowledged.

References

Abe, S., 1997. A note on the q-deformation-theoretic aspect of the generalized entropies in nonextensive physics. Physics Letters A 224, 326–330.

Aczél, J., 1984. Measuring information beyond communication theory – why some generalized information measures may be useful, others not. Aequationes Mathematicae 27, 1–19.

Aczél, J., Daróczy, Z., 1975. On Measures of Information and their Characterizations. Academic Press, London.

Aczél, J., et al., 2000. Functional equations in the behavioral sciences. Mathematica Japonica 52, 469–512.

Baumgärtner, S., 2002. Diversity as a potential for surprise. An information theoretic measure of effective product diversity. Working Paper, University of Heidelberg, available at: http://www.eco.uni-heidelberg.de/ng-oeoe/research/papers/EffDiv.pdf.

Baumgärtner, S., March 2006. Measuring the diversity of what? And for what purpose? A conceptual comparison of ecological and economic biodiversity indices. Working Paper, available at SSRN: http://ssrn.com/abstract=894782.

Ben-Bassat, M., Raviv, J., 1978. Rényi's entropy and the probability of error. IEEE Transactions on Information Theory IT-24, 324–331.

Beran, H., 1999. A generalization and optimization of a measure of diversity. In: U. Leopold-Wildburger, G. Feichtinger, K.P. Kistner (Ed.), Modelling and Decisions in Economics. Physica, Heidelberg, pp. 119–137.

Borges, E., Roditi, I., 1998. A family of nonextensive entropies. Physics Letters A 246, 399–402.

Debreu, G., 1952. Definite and semidefinite quadratic forms. Econometrica 20 (2), 295–300.

Hannah, L., Kay, J., 1977. Concentration in modern industry: Theory, measurement and the U.K. experience. The Macmillan Press, London.

Hardy, G., et al., 1934. Inequalities, 2nd Edition. Paperback Edition, reprinted 1999. Cambridge University Press, Cambridge.

Havrda, J., Charvát, F., 1967. Quantification method of classification processes: The concept of structural α-entropy. Kybernetika 3, 30–35.

Hilderman, R., Hamilton, H., 2001. Knowledge discovery and measures of interest. Kluwer Academic Publishers, Dordrecht.

Hill, M., 1973. Diversity and evenness: A unifying notation and its consequences. Ecology 54, 427–431.

Hoffmann, S., Hoffmann, A., 2006. True diversities: A comment on Lou Jost's " Entropy and diversity". Working Paper, Otto-von-Guericke-University Magdeburg, available at: http://www.ww.ovgu.de/vwl3/02_people/03_research/content_008.htm.

Jost, L., 2006. Entropy and diversity. Oikos 113, 363–374.

Kaniadakis, G., Scarfone, A., 2002. A new one-parameter deformation of the exponential function. Physica A 305, 69–75.

Kaniadakis, G., et al., 2005. Two-parameter deformations of logarithm, exponential, and entropy: A consistent framework for generalized statistical mechanics. Physical Review E 71, 0461281–04612812.

Kapur, J., 1994. Measures of information and their applications. Wiley Eastern Limited.

Keylock, C., 2005. Simpson diversity and the Shannon-Wiener index as special cases of a generalized entropy. Oikos 109, 203–206.

Kolmogorov, A., 1930. Sur la notion de la moyenne. Atti della Accademia Nazionale dei Lineci 12, 388–391.

Laakso, M., Taagepera, R., 1979. "Effective" number of parties – a measure with application to West Europe. Comparative Political Studies 12 (1), 3–27.

Mayoral, M., 1998. Rényi's entropy as an index of diversity in simple-stage cluster sampling. Journal of Information Sciences 105, 101–114.

Nagumo, M., 1930. Über eine Klasse von Mittelwerten. Japanese Journal of Mathematics 7, 71–79.

Nehring, K., Puppe, C., 2002. A theory of diversity. Econometrica 70, 1155–1198.

Patil, G., Taillie, C., 1982. Diversity as a concept and its measurement. Journal of the American Statistical Association 77 (379), 548–561.

Pielou, E., 1975. Ecological Diversity. John Wiley & Sons, New York.

Rao, C., 1982. Rao's axiomatization of diversity measures. In: Encyclopedia of Statistical Sciences. John Wiley & Sons, pp. 615–617.

Rényi, A., 1961. On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1. University of California Press, Berkeley, pp. 547–561.

Ricotta, C., 2003. On parametric evenness measures. Journal of theoretical biology 222, 189–197.

Shannon, C., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379–423.

Sharma, B., Mittal, D., 1975. New non-additive measures of entropy for discrete probability distributions. The Journal of Mathematical Sciences (Delhi) 10, 28–40.

Sharma, B., Mittal, D., 1977. New non-aditive measures of relative information. Journal of Combinatorics Information & System Sciences 2 (4), 122–132.

Sharma, B., Taneja, I., 1975. Entropy of type (α, β) and other generalized measures in information theory. Metrika 22, 205–215.

Taneja, I., 1989. On generalized information measures and their applications. In: Hawkes, P. (Ed.), Advances in Electronics and Electron Physics. Vol. 76. Academic Press, pp. 327–413.

Theil, H., 1967. Economics and Information Theory. North-Holland, Amsterdam.

Tsallis, C., 1988. Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics 52 (1/2), 479–487.

Tsallis, C., 1994. What are the numbers that experiments provide? Química nova 17 (6), 468–471.

von Neumann, J., 1927. Thermodynamik quantenmechanischer Gesamtheiten. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse 3, 273–291.

Weitzman, M., 1992. On diversity. Quarterly Journal of Economics 107, 363–405.