OTTO-VON-GUERICKE-UNIVERSITY MAGDEBURG
FACULTY OF ECONOMICS AND MANAGEMENT

Generalized distribution based diversity measurement: Survey and unification

Sönke Hoffmann

FEMM Working Paper No. 23, July 2008

F E M M

Faculty of Economics and Management Magdeburg

Working Paper Series

Otto-von-Guericke-University Magdeburg
Faculty of Economics and Management
P.O. Box 4120
39016 Magdeburg, Germany
http://www.ww.uni-magdeburg.de/

Generalized distribution based diversity measurement: Survey and unification

Februar 2007 / Rev. August 2008

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 diversity. The presents paper surveys those depending on the distribution of abundances among a given set of categories. Characteristic properties of the measures are generalized and a unifying notation is derived. It is argued that such unification 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.

Key words: Diversity measurement; Generalization; Non–additivity; Concavity; Numbers equivalence
JEL classification: Q20, C65

1 Introduction

The measurement of diversity is, by itself, a very "diverse" issue. Approaches to quantify this phenomenon vary much and the contexts to which diversity measures are applied are manifold (Junge, 1994). It is not uncommon to observe an increasing diversity according to one popular measure and a decreasing one *ceteris paribus* according to another [1] . This clearly shows that the choice between measures is all but arbitrary and characteristic properties of a specific diversity index must be discussed very carefully. Unfortunately there is not much consensus in the properties which diversity measures *should* satisfy in a given situation. Especially normative discussions between the social and natural sciences

Email address: `soenke.hoffmann@ww.uni-magdeburg.de` (Sönke Hoffmann).
[1] For a numeric example see Tóthmérész (1998), 149.

have revealed a fundamentally different understanding of diversity, which may originate in different philosophies (Baumgärtner, 2007). But even within a single science, people obviously think differently about what diversity is and how it should be measured. An *intrinsic* subjectivity of diversity explains why a multitude of diversity measures is likely to continue to exist and why there can be no "true" best–of–all measure satisfying everything for everybody (cf. Hoffmann and Hoffmann 2008). The remaining question is: How can we account for individual requirements in the choice of available measures?

The most obvious but most laborious way is to analyze each of the available indices qualitatively and to judge whether the characteristic properties satisfy the given needs. Regarding the large number of available indices this approach is not very promising. Much more efficient is the reverse way, namely to start with some basic requirement which is imposed on an appropriate generalized model of diversity in order to narrow down admissible choices. Nehring and Puppe (2002), for example, provide a general theory of diversity in terms of the qualitative disparity of attributes realized by n given objects. In subsequent papers the authors show that different properties imposed on their model lead to more specific measures (Nehring and Puppe, 2003, 2004). In this way the so–called "multi–attribute model" can be applied quite flexibly, according to individual perceptions of what diversity should be. However, diversity measures are also required to be *distribution based* sometimes, i.e. they must depend on the distribution of some quantity over the n given objects (e.g. Pielou 1975) [2]. While the multi–attribute approach seems to establish as the most general model of disparity based diversity, a unifying model of distribution based diversity just as general and flexible is still missing.

The present paper tries to complement Nehring and Puppe's (2002) theory and focuses on the generalized measurement of distribution based diversity. First it introduces the reader to the general concept of distribution based measures and gives some example applications from different sciences (Section 2). In the preceding Section 3 commonly used models of distribution based diversity are surveyed, which are then unified in Section 4. It turns out, that this "new" generalized diversity measure exists in information theory, ready to use, for more than thirty years (Sharma and Mittal, 1975). Nevertheless, the "re–discovered" Sharma–Mittal formalism not only recovers all surveyed measures by simple parameter variation but also can be made more specific according to very individual requirements. Section 5 concludes and critically gives an outlook.

[2] Note that ecologists also use attribute based diversity measures (e.g. Faith 1994) and economists sometimes use distribution based diversity measures (e.g. Beran 1999). This is, however, not the normal case. In economics distribution based measures are much more often used as measures of inequality (Atkinson, 1983) or concentration (Hannah and Kay, 1977). As we will see later many developments in the measurement of distribution based diversity can also be traced back to statistical mechanics and information theory.

2 The basic concept of distribution based measures

A distribution based measure is an *average quantity of n given categories* which is defined in terms of:

(1) the *categories*: Of what do we want to measure an average quantity?
(2) the *elementary quantity*: What kind of quantity of each category is to be averaged?
(3) the *average*: How is the desired quantity finally calculated?

All three questions can be answered very differently. In fact, the class of distribution based measures ranges across a large number of disciplines and applications. In this section the basic notation is introduced and some economic as well as non–economic answers to the above questions are presented.

2.1 Defining the categories

2.1.1 Notation

Let Z denote a set of some given and well–defined individuals such as the set of living objects in an ecosystem or the set of dollars representing the national income of some country. Formally, the set of n categories is a partition on Z such that $Z = \bigcup_{i=1}^{n} z_i$ and $z_i \cap z_j = \varnothing$ for all $i \neq j$. Let $h : 2^Z \to \mathbb{N}$, $h_i \mapsto h(z_i) = \#z_i$ then $a = (h_i)_{i=1...n}$ and $p = a/\#Z$ denote the discrete distribution of absolute and relative abundances among the categories, respectively. Further let $\mathbb{R}_+ = (0, \infty)$ and let $\mathscr{C}$ be the set of convex sets on $\mathbb{R}_+^n$ then $\mathscr{P} \subset \mathscr{C}$ is the n–unit simplex of reals such that for all $p \in \mathscr{P}$, $0 < p_i \leq 1$ and $\sum_i p_i = 1$. Two different distributions on different sets of individuals are written $p, q \in \mathscr{P}$, where $p \times q = \left(p_i q_j\right)_{i=1...n, j=1...m}$ is the joint distribution.

2.1.2 Examples

In industrial economics the categories are often defined as the firms of an industry each producing a share p_i of overall output Z and an average of all p_i is then taken as an indicator of the "concentration" in that industry (e.g. Hannah and Kay 1977). Using households as categories, each receiving a certain share of the overall income, slightly different averages are also used to measure (income) "inequality" (e.g. Atkinson 1983). Ecologist, on the other hand, usually define categories as biological species of which a relative abundance p_i of individuals can be observed in an ecosystem or a geographical region. In this case the "evenness" of individuals among species is considered as an indicator of "ecological diversity" (Ricotta, 2003). Similar diversity categories include but are not limited to a society's social groups (Harrison and Peng, 2006), a county's types of television channels (Aslama et al., 2004), an urban district's dwelling types (Les and Maher, 1998), a firm's businesses (Pitts and Hopkins, 1982), or a portfolio's financial assets (Woerheide and Persson, 1993), for which an average elementary quantity is often used as a measure for social–, TV–programme–, housing–, firm–, or portfolio–"diversity", respectively. These examples illustrate that the definition of categories is much influenced by the individual measurement context and can not be generally defined. For the sake of interdisciplinarity and generality we stay with the formal representation above and leave the question "Of what? (do we measure an average quantity)" intentionally open.

2.2 Defining the elementary quantity

Given a proper set of categories the average raw share p_i often does not describe precisely enough what is meant to be measured. These cases require some continuous and strictly monotonic transformation $v_i \mapsto \tau(p_i)$ which determines the classes' *elementary quantity* according to the most important characteristics of the phenonmenon under consideration.

2.2.1 Examples

Economic inequality and concentration measures, for example, use different definitions of shares. While a concentration share is usually defined as the output share p_i of firm i, the (income) inequality share is defined as the ratio of absolute individual income and the arithmetic mean of all incomes such that $v_i^{\text{ieq}} \mapsto \tau^{\text{ieq}}(p_i) = np_i$. Another example can be found in ecological statistics where p_i is employed as a variable of the "rarity" of a species i in a set of species (Patil and Taillie, 1982). Clearly, the value of p_i can be no reasonable rarity indicator due to the inverse behaviour of a classes relative abundance and what is commonly known as "rarity". Transformation $v_i^{\text{een}} \mapsto \tau^{\text{een}}(p_i) = 1/p_i$, also known as the *elementary equivalent number* (of equal sized classes) is more appropriate. A third example has established in classical information theory. Here, p_i is used as the probability of an event i of a discrete random variable having n possible outcomes. The *elementary information* is defined as the minimum "effort", measured in the number of binary digits, which is necessary to transmit the observed variable's realization from a source to a recipient (Rényi, 1966). It is defined as $v_i^{\text{loa}} \mapsto \tau^{\text{loa}}(p_i) := c\ln(1/p_i)$, $c \neq 0$ [3]. The choice of transformation τ^{loa} is motivated by the assumption that the "information obtained from the happening of two independent events is the sum of the informations yielded by the

[3] From the perspective of information theory the choice $c = 1/\ln(2)$ is most common to have "bits" as measurement unit.

two single events" (Aczél and Daróczy 1975, p. 3). This so–called log–additivity can be characterized using one of four well–known functional equations [4].

2.2.2 The Cauchy equations

Proposition 1 *For non–constant and continuous functions f and $x, y \in (0,\infty); c \neq 0$*

$$
\begin{aligned}
f(x+y) &= f(x)+f(y) &\Leftrightarrow\quad & f(x) = cx & \text{(lia)} \quad (1)\\
f(xy) &= f(x)+f(y) &\Leftrightarrow\quad & f(x) = c\ln(x) & \text{(loa)} \quad (2)\\
f(x+y) &= f(x)f(y) &\Leftrightarrow\quad & f(x) = e^{cx} & \text{(emu)} \quad (3)\\
f(xy) &= f(x)f(y) &\Leftrightarrow\quad & f(x) = x^{c} & \text{(pmu)} \quad (4)
\end{aligned}
$$

Proof of Proposition 1. These are classics derived from works of Cauchy (1821) and solutions can be found in every standard textbook on functional equations such as Aczél (1966). Aczél et al. (2000) provide a summary of proofs. ■

Functions satisfying (1), (2), (3), (4) are called *lin–additive (*"lia"*), log–additive (*"loa"*), exp–multiplicative (*"emu"*)* and *pow–multiplicative (*"pmu"*)*, respectively. The elementary information v_i^{loa}, for example, can be characterized by log–additivity (2) under restricted domain.

Corollary 2 (Elementary information) *Let $p, q \in [0,1]$ and f be a continuous and non–constant function, then for $c \neq 0$*

$$f(pq) = f(p)+f(q) \Leftrightarrow f(p) = c\ln\left(\frac{1}{p}\right). \tag{5}$$

Proof of Corollary 2. Let $p = \exp(-x)$ and $q = \exp(-y)$ in (5) then $g(x) := f(\exp(-x))$ gives $g(x+y) = g(x)+g(y)$ having the most general non–constant solution $g(x) = cx$. Thus $f(p) = cx = c\ln\left(\frac{1}{p}\right)$. ■

Log–additivity of shares is, however, not generally desirable. The sense and non–sense of some characteristic property clearly depends on the given application context and the phenomenon under consideration. To give an economic counterexample, Cowell (1980), p. 149 emphasizes for the measurement of income inequality that "there is no compelling reason why [log–additivity] should hold". This implies that the standard logarithm is no resonable transformation of income shares.

[4] In the following we are only interested in the most general non–constant and continuous solutions of a functional equation. For convenience we may omit the words non–constant and continuous.

Table 1
Degree–deformed logarithms, exponentials and linear functions in two equivalent representations.

Name	Degree δ deformation	Degree $d=\frac{\delta}{c}$ deformation
log–dfm. lin.	$\mathrm{lin}_\delta^{\log}(x)=\begin{cases}\frac{\log\left(\frac{\delta}{c}x+1\right)}{\delta} & ,\delta\neq 0\\ x & ,\delta=0\end{cases}$	$\mathrm{lin}_d^{\log}(x)=\begin{cases}\frac{\ln(dx+1)}{d} & ,d\neq 0\\ x & ,d=0\end{cases}$
exp–dfm. lin.	$\mathrm{lin}_\delta^{\exp}(x)=\begin{cases}c\,\frac{\exp(\delta x)-1}{\delta} & ,\delta\neq 0\\ x & ,\delta=0\end{cases}$	$\mathrm{lin}_d^{\exp}(x)=\begin{cases}\frac{e^{dx}-1}{d} & ,d\neq 0\\ x & ,d=0\end{cases}$
δ–dfm. log.	$\log_\delta(x)=\begin{cases}c\,\frac{x^\delta-1}{\delta} & ,\delta\neq 0\\ \log(x) & ,\delta=0\end{cases}$	$\log_d(x)=\begin{cases}\frac{x^{cd}-1}{d} & ,d\neq 0\\ \log(x) & ,d=0\end{cases}$
δ–dfm. exp.	$\exp_\delta(x)=\begin{cases}\left(\frac{\delta}{c}x+1\right)^{\frac{1}{\delta}} & ,\delta\neq 0\\ \exp(x) & ,\delta=0\end{cases}$	$\exp_d(x)=\begin{cases}(dx+1)^{\frac{1}{cd}} & ,d\neq 0\\ \exp(x) & ,d=0\end{cases}$

2.2.3 Generalized deformed transformations

In order to meet a larger range of individual requirements we have to use more general classes of continuous and strictly monotonic (and not necessarily log–additive) transformations. A promising way is to consider parameter–deformed functions that can be characterized by some functional equation similar to the one in Corollary 2. Here our focus is on deformed linear and logarithmic/exponential functions, which will also prove to be useful for generating mean values in subsequent sections of the present paper.

Let the common Napier logaritm be written $\log(x) := c\ln(x)$ where $\ln(x)$ denotes the natural logarithm and $c := 1/\ln(b)$ such that $b=\exp(1/c)$, $b\in\mathbb{R}_+\backslash\{1\}$ for all $c\neq 0$ is the basis of the Napier logarithm. Similarly $\exp(x) := e^{\frac{x}{c}}$ is inverse of the Napier Logarithm where e is the Euler Number. Then linear, logarithmic and exponential functions can be deformed by degree δ as depicted in the second column of Table 1. These functions will play a major role throughout this work. We shall further define $d := \delta/c$ for a nicer representation of some properties to be discussed. This equivalent notation is given in the right column of Table 1. Note that all degree zero deformations are limits of the corresponding deformation function for $\delta\to 0$ (or $d\to 0$). Further we can easily see that $\mathrm{lin}_\delta^{\log}(x)=\left(\mathrm{lin}_\delta^{\exp}\right)^{-1}(x)$, $\log_\delta(x)=\exp_\delta^{-1}(x)$, $\log\left(\exp_\delta(x)\right)=\mathrm{lin}_\delta^{\log}(x)$ and $\log_\delta\left(\exp(x)\right)=\mathrm{lin}_\delta^{\exp}(x)$. Deformed linear functions are characterized by a Cauchy–type functional equation, called *non–lin–additivity of degree d*.

Proposition 3 (Non–lin–additivity of degree d) *Let $x, y \in (0,\infty)$ and f be a continuous and non–constant function, then*

$$f(x+y) = f(x) + f(y) + d\,f(x)f(y) \Leftrightarrow f(x) = lin_d^{exp}(cx) = lin_\delta^{exp}(cx). \tag{6}$$

Proof of Proposition 3. Necessity is easy to check. For sufficiency let $d \neq 0$ and $g(x) := d\,f(x)+1$. Then the non–lin–additivity condition is equivalent to the Cauchy exponential equation $g(x+y) = g(x)g(y)$ which has the most general solution $g(x) = e^{\delta x}$, $\delta \neq 0$ given by (3). By resubstitution one obtains $f(x) = \left(e^{cdx} - 1\right)/d = \mathrm{lin}_d^{\exp}(cx)$; $c, d \neq 0$. For $d = 0$ we have lin–additivity $f(x+y) = f(x) + f(y)$ and then by (1) the most general solution is $f(x) = cx = \lim_{d\to 0} \mathrm{lin}_d^{\exp}(cx)$, $c \neq 0$. ■

Compared to deformed linear functions deformed logarithms are more often applied and have a longer history. Euler (1779) already considered the function $\log^{\mathrm{E}}_{\theta,\vartheta}(x) = -\left(x^{-\theta} - x^{-\vartheta}\right)/(\theta - \vartheta)$ and analyzed the case $\vartheta = 0$ and $\theta = \vartheta + \varepsilon$, $\varepsilon \to 0$. The first case reduces to $\log^{\mathrm{E}}_{\theta}(x) = \left(1 - x^{-\theta}\right)/\theta$, $\theta \neq 0$ and the latter to the partial derivative $\log^{\mathrm{E}}_{\theta,\varepsilon}(x) = -\partial\left(x^{-\theta}\right)/\partial\theta = x^{-\theta}\left((x^{\varepsilon}-1)/\varepsilon\right)$. The crutial feature of both functions is that they limit to the natural logarithm as $\theta \to 0$. Proposition 1 shows that the Napier logarithm is characterized by log–additivity, thus $\log^{\mathrm{E}}_{\theta}$ cannot be log–additive for $\theta \neq 0$. Indeed, this function belongs to a family recently studied by Naudts (2002) and other physicists. Here is shown that the Euler function is characterized by a property called *non–log–additivity of degree d*.

Proposition 4 (Non–log–additivity of degree d) *Let $x, y \in (0,\infty)$ and f be a continuous and non–constant function, then*

$$f(xy) = f(x) + f(y) + d\,f(x)f(y) \Leftrightarrow f(x) = \log_d(x) = \log_\delta(x). \tag{7}$$

Proof of Proposition 4. Necessity is easy to check. For sufficiency let $d \neq 0$ and $g(x) := d\,f(x)+1$. Then the non–log–additivity condition is equivalent to the Cauchy power equation $g(xy) = g(x)g(y)$ which has the most general solution $g(x) = x^{\delta}$, $\delta \neq 0$ given by (4). By resubstitution one obtains $f(x) = \left(x^{cd} - 1\right)/d = \log_d(x)$, $d \neq 0$. For $d = 0$ we have log–additivity $f(xy) = f(x) + f(y)$ and then by (2) the most general solution is $f(x) = c\ln(x) = \log(x) = \lim_{d\to 0}\log_d(x)$, $c \neq 0$. ■

The algebraic properties of deformed logarithms are different from those of the common Napier logarithm. Two of them worth noting are

$$\log_\delta(xy) = \log_\delta(x) + \log_\delta(y)x^{\delta} \tag{8}$$

$$\log_\delta\left(\frac{1}{x}\right) = -\log_\delta(x)x^{-\delta}. \tag{9}$$

We also may restrict the domain as we did for elementary information in Proposition 2.

Corollary 5 (Non–log–additive quantity) *Let $p, q \in (0,1]$ and f be a continuous and non–constant function, then*

$$f(pq) = f(p) + f(q) + d\, f(p) f(q) \Leftrightarrow f(p) = \log_\delta \left(1/p\right). \tag{10}$$

Proof of Corollary 5. Let $p = e^{-x}$ and $q = e^{-y}$ in (10) then $g(x) := f(e^{-x})$ gives $g(x+y) = g(x) + g(y) + d\, g(x) g(y)$ which has the most general non–constant solution $g(x) = \mathrm{lin}_\delta^{\exp}(cx)$ by Proposition 6. Thus $f(p) = g(x) = \mathrm{lin}_\delta^{\exp}(cx) = \mathrm{lin}_\delta^{\exp}(c \ln(1/p)) = \log_\delta \left(1/p\right)$. ■

2.2.4 Examples

In many practical applications the deformation parameter of degree–deformed functions is normalized for the sake of a more straightforward interpretation of its actual value. Most of these applications use one of the following cases

$$\theta = -\delta \tag{11}$$
$$\beta = 1 - \delta \tag{12}$$
$$\zeta = 1 + \delta. \tag{13}$$

For example, the initial form $\log_\delta(x)$ is concave only for $\delta \le 1$. If concavity needs to be maintained for increasing deformation degree we may use substitution (11) which gives Euler's function $\log_\theta^{\mathrm{E}}(x) = \left(1 - x^{-\theta}\right)/\theta$. In non–extensive statistical mechanics, pioneered by physicist Tsallis (1988), non–concavity is considered as a serious drawback (Tsallis, 2004). Therefore it seems more convenient to use (12), i.e. $\tau_\beta^{\mathrm{T}}(x) := \log_\beta(x) = \left(x^{1-\beta} - 1\right) / \left(1 - \beta\right)$ because this deformation function is concave for all $\beta \ge 0$ and has the log–additivity limit at $\beta = 1$ instead of $\delta = 0$. The same concavity parameter normalization can be applied to standard economic decision theory where a parameter deformation of the decision maker's utility function $u(x)$ can be used to model different risk attitudes [5]. However, turning the economist's perspective back to the transformation of shares p_i we can also recall Cowell's (1980) former remark on economic inequality measures. In particular, use (13) and set $c = 1/\zeta$ then $\log_\zeta^{\mathrm{CS}}(x) := \left(x^{\zeta-1} - 1\right) / (\zeta(\zeta - 1))$

[5] Moreover, it can be shown that decision makers have a constant relative risk aversion (CRRA) $\eta_u(x) := -x u''(x)/u'(x)$ if and only if $u(x) \propto \log_\delta(x)$ and a constant absolute risk aversion (CARA) $\eta_u(x)/x$ if and only if $u(x) \propto \log_\delta\left(\exp(x)\right) = \mathrm{lin}_\delta^{\exp}(x)$. Necessity is easy to check. For sufficiency one can use the proof technique of Keeney and Raiffa (1993), p.177. The concepts of absolute and relative risk aversion were introduced by Pratt (1964).

is again a deformed (natural) logarithm with $\lim_{\zeta\to 1}\tau_\zeta^{\mathrm{CS}}(x)=\ln(x)$. To see the major benefit of this function we transform the income shares in Theil's (1967) classic distribution based inequality measure $H_1^{\mathrm{CS}}(p) = 1/n\sum_{i=1}^n np_i\ln(np_i)$ using $\tau_\zeta^{\mathrm{CS}}(x)$ instead of $\ln(x)$. Then one directly obtains the most general class of *Theil–decomposable* inequality measures $H_\zeta^{\mathrm{CS}}(p) = (n\zeta(\zeta-1))^{-1}\sum_{i=1}^n\left((np_i)^\zeta-1\right)$, which was axiomatically characterized by Cowell and Kuga (1981) and Shorrocks (1980) [6]. In other words, sacrificing log–additivity in the underlying transformation of income shares allows to maintain Theil–decomposability in distribution based inequality measurement. In economic *concentration* measurement, on the other hand, still other properties appear to be more desirable than Theil–decomposability. The logarithmic character of the family H_ζ^{CS} is "neutralized" by taking the according exponential such that $\left(\log_\zeta^{\mathrm{CS}}\right)^{-1}\left(H_\zeta^{\mathrm{CS}}(p)\right) = nH_\zeta^{\mathrm{HK}}(p)$ where $H_\zeta^{\mathrm{HK}}(p) = \left(\sum_{i=1}^n p_i^\zeta\right)^{\frac{1}{\zeta-1}}$ is the most common family of concentration indices characterized by Chakravarty and Eichhorn (1991) and intuitively used as (inverted) numbers equivalent by Hannah and Kay (1977). Still another well–known family being affected by deformed logarithms is the "equally distributed equivalent" $H_\zeta^{\mathrm{A}}(p) = 1-\left(\log_\zeta^{\mathrm{A}}\right)^{-1}\left(H_\zeta^{\mathrm{CS}}(p)\right)$ discussed by Atkinson (1970), with $\log_\zeta^{\mathrm{A}}(x)=\log_{1-\zeta}^{\mathrm{CS}}(1/x)$.

We see that deformed logarithms can be used to directly link economic inequality–, concentration– and "equally distributed equivalent"–measures. Similar findings will also help to unify different families of diversity measures in Section 3.

2.3 Defining the average

The third dimension defining distribution based measures is the operator aggregating n elementary quantities into a single number. As for the definition of elementary quantities this definition should be specified according to some desirable characteristic. Most importantly it should maintain the characteristic properties of the elementary quantity under consideration.

In accordance with Bullen (2003) we give the following basic definition.

Definition 1 (Mean value function) *Let $u, v\in\mathbb{R}^n$, $\bar{u}=(c,\dots,c)$ and $\bar{0}=(0,\dots,0)$. A mapping $M:\mathbb{R}^n\to\mathbb{R}$ is called a* mean (value function)*, written shortly $M(u)=\langle u\rangle$, if it satisfies*

(1) $\langle\bar{u}\rangle = u$ (reflexivity, idempotency)
(2) $\langle u\rangle<\langle v\rangle$ if $u_i\le v_i$ for all i and $u_i<v_i$ for some i (strict monotonicity)

[6] See these papers for details on Theil–decomposability.

(3) $\lim_{v\to\bar{0}} \langle u+v\rangle = \langle u\rangle$ *(continuity)*

Note that reflexivity and strict monotonicity imply Cauchy's (1821) *internality property* $\min_i u_i \leq \langle u\rangle \leq \max_i u_i$, which is usually accepted as a natural requirement for mean values. Bonferroni (1924) proposed to functionally "generate" means using the equation $w_1\phi(u_1)+w_2\phi(u_2)=(w_1+w_2)\phi(\langle(u_1,u_2)\rangle_\phi^w)$, where w is an arbitrary non–negative weight and ϕ is a continuous and strictly monotonic function defined over a real interval, called *generating function*[7]. $\langle(u_1,u_2)\rangle_\phi^w$ satisfies the requirements of Definition 1 most immediately due to a characterizion by Aczél (1966), p. 242. For equal weights $w_1 = w_2 = w$, we get the symmetric mean $\langle(u_1,u_2)\rangle_\phi^w = \phi^{-1}\left(\left(\phi(u_1)+\phi(u_2)\right)/2\right)$ characterized by Aczél (1948). All standard means, namely the arithmetic, harmonic, geometric, quadratic, power and exponential mean, are generated by the functions $\phi(x)=x, x^{-1}, \log(x), x^2, x^c, e^{cx}$, $c\neq 0$, respectively. Let ω denote weights not necessarily equal and satisfying $\omega_i \in [0,1]$, $\sum_i \omega_i = 1$ then generalizing the Bonferroni equation to the n–case gives the *quasilinear mean of* u[8]

$$\langle u\rangle_\phi^{\boldsymbol{\omega}} := \phi^{-1}\left(\sum_{i=1}^n \omega_i\phi(u_i)\right). \tag{14}$$

Lemma 6 (Hardy et al. (1934)) *For constants $a\neq 0$ and b*

$$\tilde{\phi}(x)=a\phi(x)+b \Leftrightarrow \langle u\rangle_{\tilde{\phi}}^{\boldsymbol{\omega}} = \langle u\rangle_\phi^{\boldsymbol{\omega}}. \tag{15}$$

This implies that a specific quasilinear mean can always be generated by a continuum of functions. Because of its general nature the quasilinear mean can serve as appropriate aggregation operator in generalized distribution based measurements.

3 Distribution based diversity

The preceding section illustrated that a number of well–known quantities, such as information, physical entropy, economic inequality or concentration are measured using distribution based concepts, which are specified by a set of categories, an elementary quantity and an aggregation operator, respectively. Particularly from the classical "ecological" point of view diversity is also an average

[7] As a tribute to Kolmogorov (1930) and Nagumo (1930) for their seminal contributions to the characterization of functionally generated means, ϕ is also called *Kolmogorov Nagumo-function* (KN-function).

[8] For an axiomatic characterization of this class of mean value functions see Wang and Jiang (2005).

quantity of some properly defined set of categories. Therefore, we define such measure as the quasilinear mean $\langle v\rangle_{\phi}^{\boldsymbol{\omega}}$.

The main characteristic of all distribution based diversity measures is to comprise some "preference" for evenness. A set of classes is said to be maximally diversified whenever all classes share the same elementary quantity and there should be minimal diversity whenever the overall quantity is accumulated in a single class. Formally, let $\check{p}=(1,0\ldots,0)$ and $\bar{p}=(n^{-1},\ldots,n^{-1})$ then a distribution based diversity measure $V(p)$ should satisfy

$$V(\check{p})\leq V(p)\leq V(\bar{p}). \tag{16}$$

From the theory of majorization it is known that (16) is satisfied if V is symmetric (i.e. invariant under permutations of the variables) and concave or quasi–concave (Marshall and Olkin, 1979). Symmetry of $\langle u\rangle_{\phi}^{\boldsymbol{\omega}}$ is given by self–weighting $\omega_i=p_i$ (cf. Aczél and Daróczy 1963b) and, thus,

$$V(p):=\langle v\rangle_{\phi}=\phi^{-1}\left(\sum_{i=1}^{n}p_i\phi(\tau\left(p_i\right))\right) \tag{17}$$

can be considered as a proper basic family of measures [9]. Alternatively, (16) is satisfied whenever V is strictly concave in the sense of Schur (1923). Schur himself gave a straightworward derivative criterion, which is used for the definition of distribution based diversity measures until today (e.g. Kreutz-Delgado and Rao 1998).

Lemma 7 (Schur–condition, Schur (1923)) *Let $f:\mathscr{P}\to\mathbb{R}$ be differentiable such that $f_{p_i}\left(p\right)=\partial f\left(p\right)/\partial p_i$ exists for all i then f is strictly Schur–concave on $\mathscr{P}$ iff f is symmetric and*

$$\left(f_{p_1}(p)-f_{p_2}(p)\right)\left(p_1-p_2\right)<0 \tag{18}$$

for any two p_1,p_2 of $p\in\mathscr{P}$.

Proposition 8 (Schur–concavity of V) *Let $\tau(x)$ be a transformation of shares and $\phi\left(y\right)$ a mean generating function, both being twice differentiable. Further let $\tilde{\phi}\left(x\right):=\phi\left(\tau\left(x\right)\right)$ and $\eta_{\tilde{\phi}}\left(x\right)=-x\tilde{\phi}''(x)/\tilde{\phi}'(x)$ then $V(p)$ is strictly Schur–concave if and only if one of the conditions in Table 2 holds (line–by–line).*

Proof of Proposition 8. Define $g\left(x\right)=\tilde{\phi}\left(x\right)+x\tilde{\phi}'\left(x\right)$. Then the Schur–condition (18) is equivalent to

$$\left(\phi^{-1}\right)'\left(\sum\nolimits_{i=1}^{n}p_i\phi(\tau\left(p_i\right))\right)\cdot\left(g\left(p_1\right)-g\left(p_2\right)\right)\cdot\left(p_1-p_2\right)<0. \tag{19}$$

For convenience we can assume $p_1-p_2>0$ and omit the case $p_1-p_2<0$ (or vice versa) because the sign of $\left(g\left(p_1\right)-g\left(p_2\right)\right)\left(p_1-p_2\right)$ only depends on the

[9] The concavity of $\langle v\rangle_{\phi}$ clearly depends on ϕ and τ and will be discussed individually.

Table 2
Conditions for strict Schur–concavity of V.

$\phi(y)$	$\tilde{\phi}(x)$	$\eta_{\tilde{\phi}}(x)$
strictly increasing	strictly decreasing, concave	$\eta_{\tilde{\phi}}(x) \leq 0$
strictly decreasing	strictly increasing, convex	$\eta_{\tilde{\phi}}(x) \leq 0$
strictly decreasing	strictly increasing, concave	$0 \leq \eta_{\tilde{\phi}}(x) < 2$
strictly decreasing	strictly decreasing, convex	$0 \leq \eta_{\tilde{\phi}}(x) < 2$
strictly increasing	strictly increasing, concave	$\eta_{\tilde{\phi}}(x) > 2$
strictly increasing	strictly decreasing, convex	$\eta_{\tilde{\phi}}(x) > 2$

sign of $g'(x)$ and not on the one of $\left(p_1 - p_2\right)$. Further, for continuous and strictly monotonic functions $f'(\cdot) \gtreqless 0 \Leftrightarrow \left(f^{-1}\right)'(\cdot) \gtreqless 0$, thus $\left(\phi^{-1}\right)'(\cdot)$ can be replaced by $\phi'(\cdot)$. Then (19) simplifies to

$$\underbrace{\phi'\left(\sum_{i=1}^{n} p_i \phi(\tau(p_i))\right)}_{A} \cdot \underbrace{\left(g(p_1) - g(p_2)\right)}_{B} < 0. \tag{20}$$

Since ϕ is strictly monotonic, $A \neq 0$ and (20) is statisfied if and only if A or B is (exclusively) negative. So we need to derive conditions for the two cases $\phi'(\cdot) > 0$, $g'(x) < 0$ and $\phi'(\cdot) < 0$, $g'(x) > 0$, or explicitly

$$\phi'(\cdot) > 0,\ g'(x) = 2\tilde{\phi}'(x) + x\tilde{\phi}''(x) < 0 \tag{21}$$
$$\phi'(\cdot) < 0,\ g'(x) = 2\tilde{\phi}'(x) + x\tilde{\phi}''(x) > 0 \tag{22}$$

Note that $\tilde{\phi}'(x) \neq 0$ because ϕ and τ are strictly monotonic by defintion. Now, assume $\phi'(\cdot) > 0$ then (21) is clearly satisfied if $\tilde{\phi}'(x) < 0, \tilde{\phi}''(x) \leq 0$ which implies $\eta_{\tilde{\phi}}(x) \leq 0$. For the remaining possibilities $\tilde{\phi}'(x) < 0, \tilde{\phi}''(x) \geq 0$ and $\tilde{\phi}'(x) > 0, \tilde{\phi}''(x) \leq 0$ (21) is satisfied only if $-2\tilde{\phi}'(x) > x\tilde{\phi}''(x)$ and $2\tilde{\phi}'(x) < -x\tilde{\phi}''(x)$, which is equivalent to $\eta_{\tilde{\phi}}(x) > 2$, respectively. Consequently, the only cases which satisfy (21) are

$$\begin{array}{lll} \tilde{\phi}'(x) < 0 & \tilde{\phi}''(x) \leq 0 & \eta_{\tilde{\phi}}(x) \leq 0 \\ \tilde{\phi}'(x) < 0 & \tilde{\phi}''(x) \geq 0 & \eta_{\tilde{\phi}}(x) > 2 \\ \tilde{\phi}'(x) > 0 & \tilde{\phi}''(x) \leq 0 & \eta_{\tilde{\phi}}(x) > 2 \end{array}$$

and similarly, for (22), $\phi'(\cdot) < 0$:

$$\begin{array}{lll} \tilde{\phi}'(x) > 0 & \tilde{\phi}''(x) \geq 0 & \eta_{\tilde{\phi}}(x) \leq 0 \\ \tilde{\phi}'(x) < 0 & \tilde{\phi}''(x) \geq 0 & \eta_{\tilde{\phi}}(x) \geq 0 \text{ and } \eta_{\tilde{\phi}}(x) < 2 \\ \tilde{\phi}'(x) > 0 & \tilde{\phi}''(x) \leq 0 & \eta_{\tilde{\phi}}(x) \geq 0 \text{ and } \eta_{\tilde{\phi}}(x) < 2 \end{array}$$

■

Lemma 9 (Equivalent representations of $\langle v \rangle_\phi$) *Every continuous, monotonic transformation g of a self–weighted quasilinear mean $\langle v \rangle_\phi$ is a self–weighted quasilinear mean. In particular,*

$$g\left(\langle v \rangle_\phi\right) = \langle g(v) \rangle_{\hat{\phi}} \tag{23}$$

where $\hat{\phi}(y) = \phi\left(g^{-1}(y)\right)$, and

$$\langle v \rangle_\phi = \tau\left(\langle p \rangle_{\tilde{\phi}}\right) \tag{24}$$

where $\tilde{\phi}(x) = \phi(\tau(x))$.

Proof of Lemma 9. Let $\hat{\phi}(y) = \phi\left(g^{-1}(y)\right)$ then $\hat{\phi}^{-1}(z) = g\left(\phi^{-1}(z)\right)$ which must be a generating function because g is invertible by supposition. Then

$$g\left(\langle v \rangle_\phi\right) = g\left(\phi^{-1}\left(\sum_{i=1}^{n} p_i \phi(\tau(p_i))\right)\right) = \hat{\phi}^{-1}\left(\sum_{i=1}^{n} p_i \hat{\phi}\left(g\left(\tau(p_i)\right)\right)\right) = \langle g(v) \rangle_{\hat{\phi}}.$$

The case (24) is proved analogously. ■

3.1 Measures generated by linear functions

Let us first consider linear generating functions

$$\phi^{\text{lin}}(x) := ax + b,\ a \neq 0, \tag{25}$$

and the elementary quantities

$$v_i^{\text{een}} \mapsto \tau^{\text{een}}(p_i) = c\frac{1}{p_i} \tag{26}$$

$$v_i^{\text{loa}} \mapsto \tau^{\text{loa}}(p_i) = \log\left(\frac{1}{p_i}\right) \tag{27}$$

$$v_i^{\text{n-loa}} \mapsto \tau^{\text{n-loa}}(p_i) = \log_\delta\left(\frac{1}{p_i}\right). \tag{28}$$

3.1.1 *Linear mean of numbers equivalent quantities*

The linear mean of v_i^{een} gives

$$\langle v^{\text{een}} \rangle_{\phi^{\text{lin}}} = n. \tag{29}$$

This nominal number of categories ("richness") is easy to grasp and especially in the ecological sciences the species richness of a habitat is still a very common diversity indicator (e.g. Tilman 1996). Another feature is its range from the minimum value of 1 in the 1–species case and to a maximum value of n in the n–species case, which is often interpreted as species–equivalent diversity value. We will come back to this property later on.

3.1.2 *Linear mean of log–additive quantities*

For the log–additive elementary quantity (27) we have

$$V^{\text{S}}(p) := \left\langle v^{\text{loa}} \right\rangle_{\phi^{\text{lin}}} = \sum_{i=1}^{n} p_i v_i^{\text{loa}} = \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i}\right) \tag{30}$$

usually called "Shannon–Wiener index" in ecological diversity measurement (e.g. Pielou 1975; Magurran 2004). There is a plethora of very different interpretations on this index due to different definitions of categories, but V^{S} serves explicitly as a diversity measure much more often in the ecological sciences than anywhere else. Note that $V^{\text{S}}(p)$ is maximal at $\log(n)$ which is usually refered to as first information measure due to Hartley (1928).

Proposition 10 *$V^S(p)$ is log–additive and concave for all $p \in \mathscr{P}$.*

Proof of Proposition 10. Log–additivity of $V^{\text{S}}(p)$ is common knowledge in information theory (see e.g. Shannon 1948). As $h(x) = x^{-x}$ is concave in $(0,1]$, $g(x) = \sum f(h(x))$ is concave in $(0,1]$ for all concave f (Hardy et al., 1934). Alternatively, let $\mathbb{H}^V$ denote the Hessian of a continuous and twice differentiable function $V : \mathscr{C} \to \mathbb{R}$ then a necessary and sufficient condition for V to be concave is $p(-\mathbb{H}^V)p' \geq 0$ (Roberts and Varberg 1973, p. 103 and Debreu 1952). Here we have $p(-\mathbb{H}^{V^S})p' = \sum_{i=1}^{n} p_i = 1 > 0$. ∎

3.1.3 *Linear mean of non–log–additive quantities*

Now let the elementary quantity be non–log–additive, i.e. quantity (28). The linear mean of this quantity is for $c \neq 0$

$$\left\langle v^{\text{n-loa}} \right\rangle_{\phi^{\text{lin}}} = \sum_{i=1}^{n} p_i v_i^{\text{n-loa}} = \sum_{i=1}^{n} p_i \log_\delta\left(\frac{1}{p_i}\right) = \begin{cases} \frac{c}{\delta}\left(\sum_{i=1}^{n} p_i^{1-\delta} - 1\right) & , \delta \neq 0 \\ V^{\text{S}}(p) & , \delta = 0 \end{cases} . \quad (31)$$

Obviously this measure must be maximal at $\log_\delta(n)$, which can be called the degree–deformed Hartley–measure. Again, parameter substitutions may result in more convenient forms of (31). Without explicit reference Patil and Taillie (1982) use (11) and the Euler function $\log_\theta^{\text{E}}(x)$, $c = 1$ as underlying transformation of shares. They extensively discuss $V_\theta^{\text{PT}}(p) := \sum_{i=1}^{n} p_i \log_\theta^{\text{E}}(x)\left(1/p_i\right)$ as generalized diversity measure, which is, indeed, applied to the measurement of ecological diversity until today (e.g. Keylock 2005). An earlier reference was given by information theorists Havrda and Charvát (1967). However, this measure became most popular with the development of non–extensive statistical mechanics in the late 1980s (Tsallis, 1988). Here parameter substitution (12) and notation

$$V_\beta^{\text{T}}(p) := \sum_{i=1}^{n} p_i \log_\beta^{\text{T}}\left(\frac{1}{p_i}\right) = \begin{cases} \frac{c}{1-\beta}\left(\sum_{i=1}^{n} p_i^{\beta} - 1\right) & , \beta \neq 1 \\ V^{\text{S}}(p) & , \beta = 1 \end{cases} \quad (32)$$

is preferred which we will refer to as *Tsallis diversity*[10] .

Proposition 11 *$V_\beta^{T}(p)$ is non–log–additive for all β and concave for $\beta \geq 0$.*

Proof of Proposition 11. We show that $\left\langle v^{\text{n-loa}} \right\rangle_{\phi^{\text{lin}}}$ is non–log–additive for all δ, thus, $V_\beta^{\text{T}}(p)$ must be non–log additive for all β. Let $v_{p\times q}^{\text{n-loa}} = \left(\log_\delta\left(p_i^{-1} q_j^{-1}\right)\right)_{i=1\ldots n,\, j=1\ldots m}$ then using (8) we have for all δ

[10] In Henri Theil's view of economic inequality measurement V_β^{T} can be interpreted as decomposability–preserving generalized income *equality* (diversity) measure. More specifically, Theil (1967), p. 91 assumes that inequality is maximum equality minus equality, or formally, for the Shannon–Index as underlying equality measure $H^{\text{Th}}(p) = \ln(n) - V^{\text{S}}(p)$, $c = 1$. Since V_β^{T} is equivalent to V^{S} having degree–deformed income shares it is quite meaningful to observe that $H_\beta^{\text{CS}}(p) = \log_\beta^{\text{T}}(n) - V_\beta^{\text{T}}(p)$, $c = \beta n^{1-\beta}, \beta \neq 0$. This illustrates that the underlying equality measure in the most general class of Theil–decomposable inequality measures is a linear mean of an elementary quantity being characterized by non–log–additivity (7).

$$
\begin{aligned}
\left\langle v^{\text{n-loa}}_{p\times q}\right\rangle_{\phi^{\text{lin}}} &= \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \log_\delta\left(\frac{1}{p_i}\frac{1}{q_j}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \left(\log_\delta\left(\frac{1}{p_i}\right) + \log_\delta\left(\frac{1}{q_j}\right) p_i^{-\delta}\right) \\
&= \sum_{i=1}^{n} p_i \log_\delta\left(\frac{1}{p_i}\right)\sum_{j=1}^{m} q_j + \sum_{j=1}^{m} q_j \log_\delta\left(\frac{1}{q_j}\right)\sum_{i=1}^{n} p_i^{1-\delta} \\
&= \sum_{i=1}^{n} p_i \log_\delta\left(\frac{1}{p_i}\right) + \sum_{j=1}^{m} q_j \log_\delta\left(\frac{1}{q_j}\right)\left(\frac{c}{\delta}\sum_{i=1}^{n} p_i c\frac{p_i^{-\delta}-1}{\delta} + 1\right) \\
&= \left\langle v^{\text{n-loa}}_{p}\right\rangle_{\phi^{\text{lin}}} + \left\langle v^{\text{n-loa}}_{q}\right\rangle_{\phi^{\text{lin}}}\left(d\left\langle v^{\text{n-loa}}_{p}\right\rangle_{\phi^{\text{lin}}} + 1\right) \\
&= \left\langle v^{\text{n-loa}}_{p}\right\rangle_{\phi^{\text{lin}}} + \left\langle v^{\text{n-loa}}_{q}\right\rangle_{\phi^{\text{lin}}} + d\left\langle v^{\text{n-loa}}_{p}\right\rangle_{\phi^{\text{lin}}}\left\langle v^{\text{n-loa}}_{q}\right\rangle_{\phi^{\text{lin}}}.
\end{aligned}
$$

Concavity: For $p \in \mathscr{C}$ we have $p(-\mathbb{H}^{V^{\mathrm{T}}_{\beta}})p' = \beta\sum_{i=1}^{n} p_i^{\beta} \geq 0$ for all $\beta \geq 0$. ■

Conclusion 1 (Linear means) *Linear means $\langle v\rangle_{\phi^{lin}}$ maintain numbers equivalence, log–additivity and non–log–additivity of the elementary quantity v.*

3.2 Measures generated by non–linear functions

A natural question arising from the previous section is whether Conclusion 1, or parts of it, hold only for linear functions. The objective is to find the most general classes of generating functions ϕ maintaining numbers equivalence, log–additivity and non–log–additivity, respectively. As we have seen before, these properties characterize deformed linear or logarithmic elementary quantities. In fact, the only way to obtain a more general class of measures, satisfying these properties is to generalize the mean generating function. This deformation has other consequences on the final measure than deforming transformation functions τ. For that reason we introduce a new parameter γ and call it the *deformation degree of the mean generating function* or, according to most of the present literature, the *order*. To stay consistent with parameter substitution (12) we should also define

$$
\alpha = 1 - \gamma. \tag{33}
$$

We start with log–additive quantities (27) this time.

3.2.1 Non–linear mean of log–additive quantities

Definition 2 (Rényi-generating function) *Let $a \neq 0$ and b be arbitrary constants, then*

$$\phi_{\gamma}^{R}(x) := \begin{cases} a\exp(\gamma x)+b & ,\ \gamma\neq 0 \\ ax+b & ,\ \gamma=0 \end{cases}. \tag{34}$$

Note that $\phi_{\gamma}^{\mathrm{R}}(x) = \mathrm{lin}_{\gamma}^{\exp}(x)$ for $a = -b = \frac{c}{\gamma}$, $\gamma \neq 0$.

Proposition 12 (Log–additivity-preserving means) *Let v_i be log–additive, then*

$$\langle v\rangle_{\phi} \text{ is log–additive} \Leftrightarrow \phi(x) = \phi_{\gamma}^{R}(x).$$

Proof of Proposition 12. Rényi (1961) already proved this for conditional (relative) information measures and incomplete distributions. Here I use a related prove technique for the sufficiency part which can also be found in Hardy et al. (1934).

(1) Necessity: Let $\gamma \neq 0$ then $\left\langle v_{p\times q}^{\mathrm{loa}}\right\rangle_{\phi_{\gamma}^{\mathrm{R}}} = \frac{1}{\gamma}\log\left(\sum_{i=1}^{n}\sum_{j=1}^{m} p_i^{1-\gamma} q_j^{1-\gamma}\right) = \frac{1}{\gamma}\log\left(\sum_{i=1}^{n} p_i^{1-\gamma}\right) + \frac{1}{\gamma}\log\left(\sum_{j=1}^{m} q_j^{1-\gamma}\right)$ which is the log–additivity condition for $\left\langle v^{\mathrm{loa}}\right\rangle_{\phi_{\gamma}^{\mathrm{R}}}$. As $\lim_{\gamma\to 0}\left\langle v^{\mathrm{loa}}\right\rangle_{\phi_{\gamma}^{\mathrm{R}}} = V^{\mathrm{S}}(p)$ log–additivity is given by Proposition 10.

(2) Sufficiency: Let $\langle v\rangle_{\phi}$ be log–additive then

$$\phi^{-1}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \phi\left(\log\left(p_i^{-1} q_j^{-1}\right)\right)\right) = \phi^{-1}\left(\sum_{i=1}^{n} p_i \phi\left(\log\left(p_i^{-1}\right)\right)\right) + \phi^{-1}\left(\sum_{j=1}^{m} q_j \phi\left(\log\left(q_j^{-1}\right)\right)\right).$$

Recalling Lemma 9 we define $\tilde{\phi}(x) := \phi(-\log(x))$ viz. $\phi^{-1}(z) = -\log\left(\tilde{\phi}^{-1}(z)\right)$. Then the log–additivity condition for $q_j := \frac{1}{m}$, $m \geq 1$

$$-\log\left(\tilde{\phi}^{-1}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} p_i q_j \tilde{\phi}\left(p_i \frac{1}{m}\right)\right)\right) = -\log\left(\tilde{\phi}^{-1}\left(\sum_{i=1}^{n} p_i \tilde{\phi}(p_i)\right)\right) - \log\left(\tilde{\phi}^{-1}\left(\sum_{j=1}^{m} \frac{1}{m}\tilde{\phi}\left(\frac{1}{m}\right)\right)\right)$$

which is equivalent to the multiplicativity condition

$$\tilde{\phi}^{-1}\left(\sum_{i=1}^{n} p_i \tilde{\phi}\left(p_i \frac{1}{m}\right)\right) = \tilde{\phi}^{-1}\left(\sum_{i=1}^{n} p_i \tilde{\phi}\left(p_i\right)\right) \cdot \frac{1}{m}.$$

Resubstitution of $\tilde{\phi}^{-1}(z) = \exp\left(-\phi^{-1}(z)\right)$ gives

$$\phi^{-1}\left(\sum_{i=1}^{n} p_i \phi\left(\log\left(p_i^{-1}\right) + y\right)\right) = \phi^{-1}\left(\sum_{i=1}^{n} p_i \phi\left(\log\left(p_i^{-1}\right)\right)\right) + y \qquad (35)$$

where $y = \log(m)$ is a constant. Now let $\phi_y(x) := \phi(x+y)$ viz. $\phi^{-1}(z) = \phi_y(z) + y$ then (35) becomes $\phi_y^{-1}\left(\sum_{i=1}^{n} p_i \phi_y\left(\log\left(p_i^{-1}\right)\right)\right) = \phi^{-1}\left(\sum_{i=1}^{n} p_i \phi\left(\log\left(p_i^{-1}\right)\right)\right)$ which is true, by Lemma 6, if and only if for constants $a(y)$ and $b(y)$ depending on m

$$\phi_y(x) = \phi(x+y) = a(y)\,\phi(x) + b(y). \qquad (36)$$

In order to solve the functional equation (36) we need to note that the role of x as a variable and the one of y as a constant can be interchanged since the beginning of the proof without changing solutions. Neither does the assumption $b(y) = \phi(y)$ as we can set—again by Lemma 6—the root and the slope of ϕ without changing the resulting mean value. Under these assumptions (36) can be written as $(a(y)-1)/\phi(y) = (a(x)-1)/\phi(x)$ where one of the terms on the left or right side is a constant. If we define $d := (a(x)-1)/\phi(x)$ then we can substitute $a(y) = d\phi(y) + 1$ into (36) and obtain the non–lin–aditivity condition $\phi(x+y) = \phi(x) + \phi(y) + d\phi(x)\phi(y)$ which has the most general solution $\mathrm{lin}_\gamma^{\exp}(x)$ by Proposition 3. Finally applying Lemma 6 gives $\phi(x) = \phi_\gamma^{\mathrm{R}}(x)$.

■

Explicitly, we have

$$\begin{aligned}\left\langle v^{\mathrm{loa}}\right\rangle_{\phi_\gamma^{\mathrm{R}}} &= \left(\phi_\gamma^{\mathrm{R}}\right)^{-1}\left(\sum_{i=1}^{n} p_i \phi_\gamma^{\mathrm{R}}\left(\log\left(\frac{1}{p_i}\right)\right)\right) \\ &= \begin{cases} \frac{1}{\gamma}\log\left(\sum_{i=1}^{n} p_i^{1-\gamma}\right) & , \gamma \neq 0 \\ V^{\mathrm{S}}(p) & , \gamma = 0 \end{cases}\end{aligned} \qquad (37)$$

To obtain the notation proposed by Rényi (1961) we can apply substitution (33) and call

$$V_\alpha^{\mathrm{R}}(p) := \begin{cases} \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{n} p_i^\alpha\right) & , \alpha \neq 1 \\ V^{\mathrm{S}}(p) & , \alpha = 1 \end{cases} \tag{38}$$

the *Rényi diversity*. Log–additivity may (or may not) be a reason why V_α^{R} became "by far the most widely used class of diversity–based rankings of ecosystems in biology" (Gravel 2006, p. 19). However, general log–additivity comes at some "cost" in terms of lost properties, which can—depending on some deliberative normative discourse—be more severe. To give an example, the concavity of V_α^{R} is only given for a very limited parameter space [11] .

Proposition 13 *For all $p \in \mathscr{C}$ and $n \geq 2$, $V_\alpha^R(p)$ is concave only for $\alpha \in [0,1]$ and it is neither concave nor convex for all $\alpha > 2$.*

Proof of Proposition 13. Case 1: For $\alpha = 1$ we have $V^{\mathrm{S}}(p)$ which was shown to be strictly concave. For all $\alpha \neq 1$ we have $p(-\mathbb{H}^{V_\alpha^R})p' = \alpha/(1-\alpha) \geq 0$ iff $\alpha \in [0,1)$. For case 2 (and other cases) see Ben-Bassat and Raviv (1978), p. 326. ■

Using Lemma 9, (37) can be rewritten as $\langle v^{\mathrm{loa}}\rangle_{\phi_\gamma^{\mathrm{R}}} = \tau\left(\langle p\rangle_{\hat{\phi}}\right) = \tau\left(\hat{\phi}^{-1}\left(\sum_{i=1}^{n} p_i \hat{\phi}^{-1}(p_i)\right)\right)$ where $\hat{\phi}(x) = \phi_\gamma^{\mathrm{R}}(\tau(x)) = a\exp\left(\gamma\log\left(\frac{1}{x}\right)\right) + b = ax^{-\gamma} + b = \log_\gamma(1/x)$ for $a = -b = \frac{c}{\gamma}$, $\gamma \neq 0$. Then $\hat{\phi}^{-1}(x) = 1/\exp_\gamma(x)$, thus, $\langle v^{\mathrm{loa}}\rangle_{\phi_\gamma^{\mathrm{R}}} = \log\left(\exp_\gamma\left(\sum_{i=1}^{n} p_i \log_\gamma\left(\frac{1}{p_i}\right)\right)\right) = \mathrm{lin}_\gamma^{\log}\left(\langle v^{\text{n-loa}}\rangle_{\phi^{\mathrm{lin}}}\right)$ or equivalently $\langle v^{\text{n-loa}}\rangle_{\phi^{\mathrm{lin}}} = \mathrm{lin}_\gamma^{\exp}\left(\langle v^{\mathrm{loa}}\rangle_{\phi_\gamma^{\mathrm{R}}}\right)$. Rényi and Tsallis diversity are monotone transformations of each other, where the transformation is a degree deformed linearity.

3.2.2 Non–linear mean of numbers equivalent quantities

Despite its intuitive range and general simplicity, the nominal number of classes (29) does not account for a classes relative quantity p_i. In this paragraph we derive more general numbers equivalent measures which do. First we need to define numbers equivalence of measures depending on p.

Definition 3 (Numbers–equivalent $\langle v\rangle_\phi$) *A distribution based measure $\langle v\rangle_\phi$ is called* numbers–equivalent*, written $\langle v\rangle_\phi^{ne}$, if*

$$1 \leq \langle v\rangle_\phi \leq n \tag{39}$$

[11] Note, however, that V_δ^{R} is still quasi–concave for all non–negative δ and therefore (16) is satisfied.

and

$$\langle v\rangle_\phi = 1 \text{ for all } \check{p} \text{ and } \langle v\rangle_\phi = n \text{ for all } \bar{p}. \tag{40}$$

Proposition 14 (Numbers–equivalent $\langle v\rangle_\phi$) *Let $\langle v\rangle_\phi$ be Schur–concave, then*

$$\tau(x) = \frac{1}{x} \Rightarrow \langle v\rangle_\phi = \langle v\rangle_\phi^{ne} \tag{41}$$

Proof of Proposition 14. The first equation of (40) is equivalent to $\phi^{-1}\left(\sum_{i=1}^n \frac{1}{n}\phi\left(\tau\left(\frac{1}{n}\right)\right)\right) = n \Leftrightarrow \tau\left(\frac{1}{n}\right) = n \Leftrightarrow \tau(x) = \frac{1}{x}$ and the second to $\phi^{-1}\left(\sum_{i=1}^n 1\phi(\tau(1))\right) = 1 \Leftrightarrow \tau(1) = 1 \Leftarrow \tau(x) = \frac{1}{x}$. Together with Schur–Concavity this implies inequality (39), which proves (41). ■

Proposition 14 implies that for $\phi(x) = \hat{\phi}\left(g(x)\right)$

$$\langle g(v)\rangle_{\hat{\phi}} = \hat{\phi}^{-1}\left(\sum_{i=1}^n p_i\hat{\phi}\left(g\ \left(\tau\left(p_i\right)\right)\right)\right) = g\left(\phi^{-1}\left(\sum_{i=1}^n p_i\phi\left(\tau\left(p_i\right)\right)\right)\right) = g\left(\langle v\rangle_\phi\right)$$

is numbers equivalent if $g(\tau(x)) = 1/x$ viz. $g\left(y\right) = 1/\tau^{-1}\left(y\right)$. For example the Tsallis diversity (31) employs $\tau(x) = \log_\delta(1/x)$ as transformation of shares, thus, $g\left(\left\langle v^{\text{n-loa}}\right\rangle_{\phi^{\text{lin}}}\right) = \exp_\delta\left(\left\langle v^{\text{n-loa}}\right\rangle_{\phi^{\text{lin}}}\right)$ must be numbers equivalent. Similarly, Rényi diversity (37) is a quasilinear mean of log–additive shares and therefore can be made numbers equivalent by $g\left(y\right) = 1/\log^{-1}(1/x) = \exp(x)$. The more obvious way to derive numbers equivalents more general than the nominal number (29) is an order γ generalization rather than a δ–deformation of $\tau(x) = 1/x$. For example, when non–log–additivity (7) is imposed on the generating function we can define the following generating function.

Definition 4 (Hill generating function) *Let $a \neq 0$ and b be arbitrary constants, then*

$$\phi_\gamma^H(x) := \begin{cases} ax^\gamma + b & , \gamma \neq 0 \\ a\log(x) + b & , \gamma = 0 \end{cases} \tag{42}$$

Note that $\phi_\gamma^{\text{H}}(x) = \log_\gamma(x)$ for $a = -b = \frac{c}{\gamma}$, $\gamma \neq 0$. The self–weighted, numbers–equivalent power mean is explicitly written

$$\begin{aligned} \langle v^{\text{een}}\rangle_{\phi_\gamma^{\text{H}}} &= \left(\phi_\gamma^{\text{H}}\right)^{-1}\left(\sum_{i=1}^n p_i\phi_\gamma^{\text{H}}\left(\frac{1}{p_i}\right)\right) \\ &= \begin{cases} c\left(\sum_{i=1}^n p_i^{1-\gamma}\right)^{\frac{1}{\gamma}} & , \gamma \neq 0 \\ c\exp\left(V^{\text{S}}(p)\right) = c\prod_{i=1}^n p_i^{-p_i} & , \gamma = 0 \end{cases}. \end{aligned} \tag{43}$$

As usual we may substitue (33) and call

$$V_{\alpha}^{\mathrm{H}}(p) := \begin{cases} c\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{\frac{1}{1-\alpha}} & , \alpha \neq 1 \\ c \exp\left(V^{\mathrm{S}}(p)\right) & , \alpha = 1 \end{cases} \tag{44}$$

Hill numbers or *Hill diversity* due to the work of Hill (1973) on ecological diversity measurement. The straightforward interpretation of this class as the numbers equivalent made it also appealing to many other disciplines, such as the measurement of industrial concentration (Hannah and Kay, 1977) or party fragmentation in a parliament (Laakso and Taagepera, 1979). However, $V_{\gamma}^{\mathrm{H}}(p)$ can only be characterized by properties stronger than numbers equivalence.

Definition 5 (Replication) *Let* $p = (p_1, p_2, \ldots, p_n)$ *then*

$$_m p := \left(\underbrace{\frac{p_1}{m}, \ldots, \frac{p_1}{m}}_{m \text{ times}}, \underbrace{\frac{p_2}{m}, \ldots, \frac{p_2}{m}}_{m \text{ times}}, \ldots, \underbrace{\frac{p_n}{m}, \ldots, \frac{p_n}{m}}_{m \text{ times}} \right)$$

is called the m*–fold replication of* p*. Similarly we will write* $_m v$ *for replicated* τ*–transformed shares.*

Definition 6 (Replication–homogeneity) $\langle v \rangle_{\phi}$ *is called replication–homogenous if*

$$m \langle v \rangle_{\phi} = \langle _m v \rangle_{\phi} . \tag{45}$$

Replication–homogeneity has numerous intuitive interpretations. In ecological diversity measurement, for example, a 2–fold replicated species–abundance relation should be exactly twice as diverse as the original one (Hill 1973, p. 430). In fact, this requirement is sufficient to characterize V_{α}^{H} among all $\langle v \rangle_{\phi}$.

Proposition 15 (Replication–homogenous means)

$$m \langle v \rangle_{\phi} = \langle _m v \rangle_{\phi} \Leftrightarrow \langle v \rangle_{\phi} = V_{\gamma}^{H}(p).$$

Proof of Proposition 15. Necessity is easy to check. For sufficiency I adopt again the proof technique of Hardy et al. (1934). From replication–homogeneity of non–numbers equivalent means follows

$$\begin{aligned} \frac{m}{\phi^{-1}\left(\sum_{i=1}^{n} p_i \phi\left(p_i\right)\right)} &= \frac{1}{\phi^{-1}\left(m \frac{p_1}{m} \phi\left(\frac{p_1}{m}\right) + m \frac{p_2}{m} \phi\left(\frac{p_2}{m}\right) + \ldots + m \frac{p_n}{m} \phi\left(\frac{p_n}{m}\right)\right)} \\ &\Leftrightarrow \langle p \rangle_{\phi} = \langle p \rangle_{\tilde{\phi}} \end{aligned} \tag{46}$$

where $\tilde{\phi}(x) := \phi\left(\frac{x}{m}\right)$ such that $\phi^{-1}(z) = \frac{1}{m}\tilde{\phi}^{-1}(z)$. Then (46) is true by Lemma 6 if and only if for two constants, depending on m

$$
\begin{aligned}
\tilde{\phi}(x) = \phi\left(\frac{x}{m}\right) &= a(m)\phi(x) + b(m) \\
\Leftrightarrow \hat{\phi}(ym) &= a(m)\hat{\phi}(y) + b(m)
\end{aligned}
\tag{47}
$$

with $y = 1/x$ and $\hat{\phi}(y) = \phi(1/y)$. As in the proof of Proposition 12 we can interchange the constant and the variable and assume $\hat{\phi}(1) = 0$ viz. $\hat{\phi}(m) = b(m)$ without changing the solutions of (47), such that $d := (a(m)-1)/\hat{\phi}(m) = (a(y)-1)/\hat{\phi}(y)$. Resubstituting $a(m) = d\hat{\phi}(m)+1$ into (47) gives $\hat{\phi}(ym) = \hat{\phi}(y) + \hat{\phi}(m) + d\hat{\phi}(m)\hat{\phi}(y)$, which has the most general solution $\hat{\phi}(y) = \log_\gamma(y)$ by Proposition 4. Finally with $\hat{\phi}(y) = \phi(1/y)$ and $y = 1/x$ we obtain

$$
\begin{aligned}
\phi(x) = \log_\gamma(x) &= \begin{cases} c\frac{x^\gamma - 1}{\gamma} & , \gamma \neq 0 \\ \log(x) & , \gamma = 0 \end{cases} \\
&= \phi_\gamma^{\mathrm{H}}(x) \text{ for } a = -b = \frac{c}{\gamma}, \gamma \neq 0
\end{aligned}
$$

■

Another characterization becomes obvious with the following Lemma.

Lemma 16 *Let $g(x) = \exp(x)$ then*

$$\langle v\rangle_\phi \text{ is log–additive} \Leftrightarrow g\left(\langle v\rangle_\phi\right) \text{ is pow–multiplicative.}$$

Proof of Lemma 16.

$$
\begin{aligned}
\left\langle v_{p\times q}\right\rangle_\phi &= \left\langle v_p\right\rangle_\phi + \left\langle v_q\right\rangle_\phi \\
\Leftrightarrow \exp\left(\left\langle v_{p\times q}\right\rangle_\phi\right) &= \exp\left(\left\langle v_p\right\rangle_\phi\right)\exp\left(\left\langle v_q\right\rangle_\phi\right) \\
\Leftrightarrow g\left(\left\langle v_{p\times q}\right\rangle_\phi\right) &= g\left(\left\langle v_p\right\rangle_\phi\right) g\left(\left\langle v_q\right\rangle_\phi\right)
\end{aligned}
$$

■

Since V_γ^{R} is already characterized as the most general log–additive $\langle v\rangle_\phi$ by Propsition 12, V_γ^{H} must be the most general pow–multiplicative $\langle v\rangle_\phi$ [12].

[12] See Aczél and Daróczy (1963a) for similar findings.

3.2.3 Non–linear mean of non–log–additive quantities

Finally we need to derive the most general class of mean generating functions ϕ maintaining non–log–additivity of v. To this end, a reformulation of Lemma 16 is useful.

Lemma 17 *Let $c \neq 0$, $g(x) = \log_\delta(\exp(x)) = lin_\delta^{exp}(x)$ then*

$$\langle v \rangle_\phi \text{ is log–additive} \Leftrightarrow g\left(\langle v \rangle_\phi\right) \text{ is non–log–additive of degree } d.$$

Proof of Lemma 17. Let $\langle v \rangle_\phi$ be log–additive and $c, \delta \neq 0$ then

$$\begin{aligned}
& \left\langle v_{p\times q} \right\rangle_\phi = \left\langle v_p \right\rangle_\phi + \left\langle v_q \right\rangle_\phi \\
\Leftrightarrow\ & \exp\left(\delta \left\langle v_{p\times q} \right\rangle_\phi\right) = \exp\left(\delta \left\langle v_p \right\rangle_\phi\right) \exp\left(\delta \left\langle v_q \right\rangle_\phi\right) \\
\Leftrightarrow\ & \exp\left(\delta \left\langle v_{p\times q} \right\rangle_\phi\right) = \exp\left(\delta \left\langle v_p \right\rangle_\phi\right) + \exp\left(\delta \left\langle v_q \right\rangle_\phi\right) \\
& \qquad + \left[\exp\left(\delta \left\langle v_p \right\rangle_\phi\right) - 1\right] \left[\exp\left(\delta \left\langle v_q \right\rangle_\phi\right) - 1\right] - 1 \\
\Leftrightarrow\ & c\frac{\exp\left(\delta \left\langle v_{p\times q} \right\rangle_\phi\right) - 1}{\delta} = c\frac{\exp\left(\delta \left\langle v_p \right\rangle_\phi\right) - 1}{\delta} + c\frac{\exp\left(\delta \left\langle v_q \right\rangle_\phi\right) - 1}{\delta} \\
& \qquad + \frac{\delta}{c} c\frac{\left[\exp\left(\delta \left\langle v_p \right\rangle_\phi\right) - 1\right]}{\delta} \cdot c\frac{\left[\exp\left(\delta \left\langle v_q \right\rangle_\phi\right) - 1\right]}{\delta} \\
\Leftrightarrow\ & g\left(\left\langle v_{p\times q} \right\rangle_\phi\right) = g\left(\left\langle v_p \right\rangle_\phi\right) + g\left(\left\langle v_q \right\rangle_\phi\right) + d\, g\left(\left\langle v_p \right\rangle_\phi\right) g\left(\left\langle v_q \right\rangle_\phi\right)
\end{aligned}$$

which is the non–log–additivity condition for the function $g(\langle v \rangle_\phi) = \mathrm{lin}_\delta^{\exp}\left(\langle v \rangle_\phi\right)$. In the $\delta \to 0$ limit g is linear and non–log–additivity reduces back to log–additivity. ■

Example 18 (Gaussian Entropy) *The Shannon–Index $V^S(p)$ is log–additive due to Proposition 10. Thus, by Lemma 17 $V^G(p) := lin_\delta^{exp}(V^S(p)) = \log_\delta(\exp(V^S(p)))$ must be non–log–additive. This index is well–known in physics as "(non–extensive) Gaussian entropy" (Frank, 2004). For an equivalent quasilinear representation of V^G we can make use of Lemma 9 such that $V^G(p) = \langle v^{n\text{-}lad} \rangle_{\phi^G}$ and $\phi^G(x) = a\, lin_\delta^{log}(y) + b$, $a \neq 0$.*

Given example 18, the following generalization appears natural.

Proposition 19 (Non–log–additivity–preserving means) *Let τ be non–log–additive of degree d, then $\langle v \rangle_\phi$ is non–log–additive of degree d if and only if*

$\phi^{SM}(x) = a\phi^*(x) + b$, $a \neq 0$ *where*

$$\phi^*(x) = \begin{cases} \left(\frac{\delta}{c}x+1\right)^{\frac{\gamma}{\delta}} & ;\delta \neq \gamma \neq 0 \\ \log\left(\frac{\delta}{c}x+1\right) & ;\delta \neq 0;\gamma = 0 \\ \exp(\gamma x) & ;\delta = 0;\gamma \neq 0 \\ x & ;\delta = \gamma \end{cases}. \tag{48}$$

Proof of Proposition 19. A proof is given in Hoffmann (2008). Here I shall present a more simple alternative. We know that the Rényi diversity $\left\langle v^{\mathrm{loa}} \right\rangle_{\phi_\gamma^{\mathrm{R}}}$ is the most general log–additive $\langle v \rangle_\phi$. Thus, by Lemma 17

$$\begin{aligned} V_{\gamma,\delta}^{\mathrm{SM}}(p) &= \mathrm{lin}_\delta^{\exp}\left(\left\langle v^{\mathrm{loa}} \right\rangle_{\phi_\gamma^{\mathrm{R}}}\right) \\ &= \log_\delta\left(\langle v^{\mathrm{een}} \rangle_{\phi_\gamma^{\mathrm{H}}}\right) = \log_\delta\left(\exp_\gamma\left(\left\langle v^{\text{n-loa}} \right\rangle_{\phi^{\mathrm{lin}}}\right)\right) \end{aligned} \tag{49}$$

is the most general non–log–additive $\langle v \rangle_\phi$. The remaining task is to find a $\tilde{\phi}$ such that (49) can be written as $\langle v \rangle_{\tilde{\phi}}$. Again, we recall Lemma 9 and obtain

$$V_{\gamma,\delta}^{\mathrm{SM}}(p) = \left\langle v^{\text{n-loa}} \right\rangle_{\tilde{\phi}}$$

where $g(x) = \mathrm{lin}_\delta^{\exp}(x)$, $\phi(x) = \phi_\gamma^{\mathrm{R}}(x) \stackrel{!}{=} \mathrm{lin}_\gamma^{\exp}(x)$, $\tau(x) = \log(1/x)$ such that

$$\begin{aligned} \tilde{\phi}(z) &= \phi\left(g^{-1}(z)\right) = \mathrm{lin}_\gamma^{\exp}\left(\mathrm{lin}_\delta^{\log}(z)\right) = \log_\gamma\left(\exp_\delta(z)\right) \\ &= \begin{cases} \log_\gamma\left(\exp_\delta(z)\right) = c\,\frac{\left(\frac{\delta}{c}z+1\right)^{\frac{\gamma}{\delta}}-1}{\gamma} & ;\delta \neq \gamma \neq 0 \\ \mathrm{lin}_\delta^{\log}(z) = \frac{\log\left(\frac{\delta}{c}z+1\right)}{\delta} & ;\delta \neq 0;\gamma = 0 \\ \mathrm{lin}_\gamma^{\exp}(z) = c\,\frac{\exp(\gamma z)-1}{\gamma} & ;\delta = 0;\gamma \neq 0 \\ z & ;\delta = \gamma \end{cases} \\ \tilde{\tau}(x) &= g(\tau(x)) = \mathrm{lin}_\delta^{\exp}\left(\log(1/x)\right) = \log_\delta(1/x). \end{aligned}$$

Finally, applying (15) gives the solution (48). Note that the $\delta = 0; \gamma \neq 0$ case recovers non–additivity of degree zero, i.e. log–additivity. ■

4 A unification and comparison framework

In the preceding section different commonly used families of distribution based diversity measures were characterized and related to each other. Now all these

Table 3
Examples of popular distribution based diversity index–families included in the Sharma–Mittal function.

α	β	Mean $\phi(x)$	Quantity $\tau(p_i)$	Measure
$\neq 1$	$\neq 1$	$\log_{1-\alpha}\left(\exp_{1-\beta}(x)\right)$	$\log_{1-\beta}(1/p_i)$	$V^{\mathrm{SM}}_{\alpha,\beta}$ (Sharma–Mittal)
$=1$	$\neq 1$	$\mathrm{lin}^{\log}_{1-\beta}(x)$	$\log_{1-\beta}(1/p_i)$	V^{G}_{β} (Gauss)
$\neq 1$	$=1$	$\mathrm{lin}^{\exp}_{1-\alpha}(x)$	$\log(1/p_i)$	V^{R}_{α} (Rényi)
$=\beta$	$\neq 1$	x	$\log_{1-\beta}(1/p_i)$	V^{T}_{β} (Tsallis)
$=\beta$	$=1$	x	$\log(1/p_i)$	V^{S} (Shannon)
$\neq 0$	$=0$	$\log_{1-\alpha}\left(\frac{x}{c}+1\right)$	$\frac{c}{p_i}-c$	$V^{\mathrm{H}}_{\alpha}(p)-c$ (Hill)
$=0$	$=0$	x	$\frac{c}{p_i}-c$	$cn-c$ (Richness)
$=0$	$\neq 0$	x	$\log_{1-\beta}(1/p_i)$	$\log_{\beta}(n)$ (Dfm. Hartley)
$=0$	$=1$	$\exp(x)$	$\log(1/p_i)$	$\log(n)$ (Hartley)

families (including their respective indices) should be unified within a single consistent framework. Recalling the generalization methods discussed so far, it seems obvious to derive a two–parameter notation, where δ represents the degree of non–additivity of the quantity and γ the deformation degree in the mean generating process. Whenever $\delta = 0$ all γ–deformed means should be log–additive and $\delta \neq 0$ indicates the degree of non–additivity of a specific γ–deformed mean. If $\gamma = 0$ the mean is either linear or logarithmic, depending on whether we use the deformed linearity or logarithm, and finally for $\delta = \gamma$ we do not distinguish between deformation of the quantity and of the aggregation operator. In fact, all these cases are unified by $\left\langle v^{\text{n-loa}}\right\rangle_{\phi^{\mathrm{SM}}}$ as

$$V^{\mathrm{SM}}_{\gamma,\delta}(p) := \left\langle v^{\text{n-loa}}\right\rangle_{\phi^{\mathrm{SM}}} = \begin{cases} \frac{c}{\delta}\left(\left(\sum_{i=1}^{n} p_i^{1-\gamma}\right)^{\frac{\delta}{\gamma}} - 1\right) & ;\delta \neq \gamma \neq 0 \\ \frac{c}{\delta}\left(\exp\left(\delta V^{\mathrm{S}}(p)\right) - 1\right) & ;\gamma = 0;\ \delta \neq 0 \\ \frac{c}{\gamma}\ln\left(\sum_{i=1}^{n} p_i^{1-\gamma}\right) & ;\gamma \neq 0;\ \delta = 0 \\ \frac{c}{\delta}\left(\sum_{i=1}^{n} p_i^{1-\delta} - 1\right) & ;\gamma = \delta \neq 0 \\ c\sum_{i=1}^{n} p_i \ln\left(\frac{1}{p_i}\right) & ;\gamma = \delta = 0 \end{cases} .$$

This two parameter generalization was introduced to information theory by Sharma and Mittal (1975, 1977). In the present context we call it the *diversity* *of order* γ *and degree* δ, or simply *Sharma–Mittal diversity*[13]. A number of well–known measures can be identified with a specific parameter tuple. Having applied substitutions (12) and (33) some discrete points in the α–β–parameter space and the corresponding measure can be recovered as listed in Table 3. As we have seen in the preceding sections most of these can be characterized by some Cauchy–type functional equation. The continuity of a characterizing property represented by a continuous parameter may help finding an adequate "degree" to which a desirable property should be satisfied. The key–property considered here is non–additivity up to some degree. Of course other properties may also be analyzed. Assume, for example, that concavity is a necessary condition in some diversity measurement context, such as stochastic portfolio theory (Fernholz 2002, 62 pp.). Then we do not need to check each of the known indices for concavity, but, instead we can also try to determine the α–β–parameter space for concavity which may directly provide an admissible family of distribution based measures. As an illustration and for the sake of completeness the following proposition is given (cf. Figure 1).

Proposition 20 (Concavity of $V^{\mathbf{SM}}_{\alpha,\beta}$) *$V^{SM}_{\alpha,\beta}$ is strictly concave for all $\beta > 1-(1-\alpha)/\alpha$.*

Proof of Proposition 20. Let $p \in \mathscr{C}$ and $r = 1\ldots n-1$ be the vertical and $s = 1\ldots n-1$ the horizontal index of the Hessian $\mathbb{H}^{V^{\mathrm{SM}}_{\alpha,\beta}}$. Further define $y(p) := \sum_{r=1}^{n} p_r^{\alpha}$, $z(p) := -\frac{\alpha}{(1-\alpha)^2} y(p)^{\frac{1-\beta}{1-\alpha}-2}$, $\mathbf{a} = \alpha\beta - \alpha^2$ and $\mathbf{b} = (\alpha-1)^2$ then

$$\mathbb{H}^{V^{\mathrm{SM}}_{\alpha,\beta}} =$$

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

and

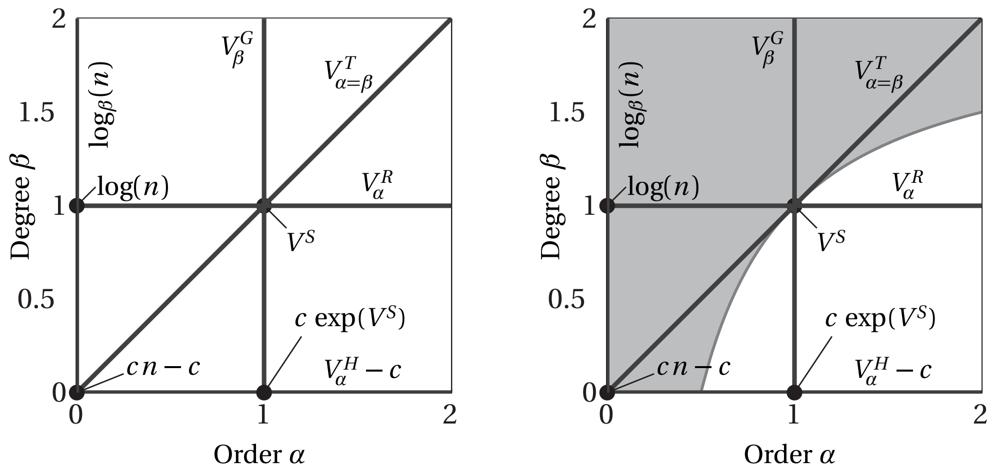
$$\begin{aligned} p\left(-\mathbb{H}^{V^{\mathrm{SM}}_{\alpha,\beta}}\right)p' &= \frac{\alpha(1-2\alpha+\alpha\beta)}{\underbrace{(\alpha-1)^2}_{>0}} \underbrace{\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{\frac{1-\beta}{1-\alpha}}}_{>0} \overset{!}{>} 0 \\ &\Leftrightarrow \alpha(1-2\alpha+\alpha\beta) > 0 \\ &\Leftrightarrow \beta > \frac{2\alpha-1}{\alpha} = 1-\frac{1-\alpha}{\alpha} \end{aligned}$$

[13] In information theory this measure is also called *information of order a and rank b* (Aczél, 1984).

■

Figure 1. Well–known distribution–based measures in the Sharma–Mittal parameter space. The black lines and dots represent the measures as a function of their parameter.

(a) Unrestricted parameter–space. (b) Concavity–parameter–space.



5 Conclusion and outlook

The preceding sections summarized mathematical characterizations for popular index–families being used in diversity measurement and provided a unification which recovers all of these families by parameter variation. As old as the unification formalism is, as unknown it seems to be in applied diversity theory. This is not very reasonable because a consistent comparison framework can significantly simplify the rational choice of measures under some individual requirements. It must be emphasized, however, that the Sharma–Mittal model does not constitute a good or admissible diversity meansure *per se*. Contexts in which distribution based measures are generally useless can be found easliy [14] . In this paper, I advocate the Sharma–Mittal measure only for its strictly formal feature to unify popular, *existing* measures (be they "good" or "bad") and to express their varying qualitative differences in terms of varying numeric values.

The Sharma–Mittal formalism clearly provides helpful insights to a large variety of diversity measures. Nevertheless its qualitative scope is still limited and partial. To put it differently, some properties may be important in some context for which there is no single point in the α–β–parameter space. These cases need generalizations different from deforming functions in the way it was discussed here. Abe (1997), for example, presents a β–deformed logarithm having

[14] See Weitzman (1992) or Nehring and Puppe (2002).

the $\beta \longleftrightarrow 1/\beta$ invariance property. His resulting entropy recovers Shannon entropy for $\beta \longrightarrow 1$ like most other generalizations. More recently, Kaniadakis and Scarfone (2002) introduced the κ–deformed–logarithm $\ln_\kappa(x) = (x^\kappa - x^{-\kappa})/2\kappa$, $\lim_{\kappa\to 0} \ln_\kappa(x) = \ln(x)$ which has the property $\ln_\kappa(1/x) = -\ln_\kappa(x)$ known from the Napier logarithm. 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 $H_{\kappa,r}(p) = -\sum p_i \ln_{\kappa,r}(p_i)$ 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) [15]. Tsallis entropy and Abe entropy are prominent special cases of these generalizations. However, analyzing these approaches in greater detail would quickly go beyond the scope of this paper and remains to future research.

Acknowledgements

I thank Joachim Weimann, 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., 1948. On mean values. Bulletin of the American Mathematical Society 54, 392–400.

Aczél, J., 1966. Lectures on Functional Equations and Their Applications. Academic Press, New York.

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., 1963a. Charakterisierung der Entropien positiver Ordnung und der Shannon'schen Entropie. Acta Mathematica Acad. Sci. Hung. 14, 95–121.

Aczél, J., Daróczy, Z., 1963b. Über verallgemeinerte quasilineare Mittelwerte, die mit Gewichtsfunktionen gebildet sind. Publicationes Mathematicae (Debrecen) 10, 171–190.

[15] The Borges/Roditi formalism is obtained by simple paramter transformation $\kappa \longleftrightarrow (\beta-\alpha)/2$ 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).

Aczél, J., Daróczy, Z., 1975. On Measures of Information and Their Characterizations. Academic Press, San Diego.

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

Aslama, M., et al., 2004. Does market–entry regulation matter? Gazette: The International Journal for Communication Studies 66, 113–132.

Atkinson, A., 1970. On the measurement of inequality. Journal of Economic Theory 2, 244–263.

Atkinson, A., 1983. The Economics of Inequality, 2nd Edition. Oxford University Press, New York.

Baumgärtner, S., 2007. Why the measurement of species diversity requires prior value judgments. In: Kontoleon, A., Pascual, U., Swanson, T. (Eds.), Biodiversity Economics. Cambridge University Press, UK, pp. 635–674.
URL SSRNeLibrary:http://ssrn.com/paper=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: Leopold-Wildburger, U., Feichtinger, G., Kistner, K. (Eds.), Modelling and Decisions in Economics. Physica, Heidelberg, pp. 119–137.

Bonferroni, C., 1924. La media esponenziale in matematica finanziaria. Annuario del Regio Istituto Superiore di Scienze Economiche e Commeciali di Bari AA23-24, 1–14.

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

Bullen, P., 2003. Handbook of Means and Their Inequalities. Kluwer, Dordrecht.

Cauchy, A., 1821. Cours d'analyse de l'École Royale Polytechnique. Debure.
URL http://mathdoc.emath.fr/cgi-bin/oeitem?id=OE_CAUCHY_2_3_P5_0

Chakravarty, S., Eichhorn, W., 1991. An axiomatic characterization of a generalized index of concentration. The Journal of Productivity Analysis 2, 103–112.

Cowell, F., 1980. Generalized entropy and the measurement of distributional change. European Economic Review 13, 147–159.

Cowell, F., Kuga, K., 1981. Additivity and the entropy concept: An axiomatic approach to inequality measurement. Journal of Economic Theory 25, 131–143.

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

Euler, L., 1779. De serie lambertina plurimisque eius insignibus proprietatibus. Acta Academiae Scientarum Imperialis Petropolitinae II, 29–51.
URL http://math.dartmouth.edu/~Eeuler/docs/originals/E532.pdf

Faith, D., 1994. Phylogenetic pattern and the quantification of organismal biodiversity. Philosophical Transactions: Biological Sciences 345, 45–58.

Fernholz, R., 2002. Stochastic Portfolio Theory. Springer, New York.

Frank, T., 2004. Complete description of a generalized ornstein–uhlenbeck process related to the nonextensive gaussian entropy. Physica A 340, 251–256.

Gravel, N., 2006. What is diversity, working Paper, CSH (Delhi).

URL http://www.csh-delhi.com/team/downloads/publiperso/surveygalway.pdf

Hannah, L., Kay, J., 1977. Concentration in Modern Industry: Theory, Measurement and the U.K. Experience. Macmillan, London.

Hardy, G., et al., 1934. Inequalities. Cambridge University Press, Cambridge.

Harrison, D., Peng, H., 2006. What is diversity and how should it be measured. In: Konrad, A., et al. (Eds.), Handbook of workplace diversity. Sage, London, pp. 191–216.

Hartley, R., 1928. Transmission of information. Bell Systems Technical Journal 7, 535–563.

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

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

Hoffmann, S., 2008. Non–uniqueness of non–extensive entropy under rényis recipe, fEMM Working Paper Series, No. 11, Otto–von–Guericke Universität Magdeburg.
URL http://www.ww.uni-magdeburg.de/fwwdeka/femm/a2008_Dateien/2008_11.pdf

Hoffmann, S., Hoffmann, A., 2008. Is there a "true" diversity? Ecological Economics 65, 213–215.

Junge, K., 1994. Diversity of ideas about diversity measurement. Scandinavian Journal of Psychology 35, 16–26.

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, New Delhi.

Keeney, R., Raiffa, H., 1993. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge University Press.

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.

Kreutz-Delgado, K., Rao, B., 1998. Measures and algorithms for best basis selection. Proceedings of the ICASSP 3, 1881–1884.

Laakso, M., Taagepera, R., 1979. "Effective" number of parties — a measure with application to west europe. Comparative Political Studies 12, 3–27.

Les, M., Maher, C., 1998. Measuring diversity: Choice in local housing markets. Geographical Analysis 30, 172–190.

Magurran, A., 2004. Measuring Biological Diversity. Blackwell Publishing, Oxford.

Marshall, A., Olkin, I., 1979. Inequalities: Theory of Majorization and Its Appli-

cations. Vol. 143 of Mathematics in Science and Engineering. Academic Press, San Diego.
Nagumo, M., 1930. Über eine klasse von mittelwerten. Japanese Journal of Mathematics 7, 71–79.
Naudts, J., 2002. Deformed exponentials and logarithms in generalized thermostatistics. Physica A 316, 323–334.
Nehring, K., Puppe, C., 2002. A theory of diversity. Econometrica 70, 1155–1198.
Nehring, K., Puppe, C., 2003. Diversity and dissimilarity in lines and hierarchies. Mathematical Social Sciences 45, 167–183.
Nehring, K., Puppe, C., 2004. Modelling phylogenetic diversity. Resource and Energy Economics 26, 205–235.
Patil, G., Taillie, C., 1982. Diversity as a concept and its measurement. Journal of the American Statistical Association 77, 548–561.
Pielou, E., 1975. Ecological Diversity. Wiley, New York.
Pitts, R., Hopkins, H., 1982. Firm diversity: Conceptualization and measurement. Academy of Management Review 7, 620–629.
Pratt, J., 1964. Risk aversion in the small and in the large. Econometrica 32, 122–136.
Rényi, A., 1961. On measures of entropy and information. In: Neyman, J. (Ed.), Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1. University of California Press, Berkeley, pp. 547–561.
Rényi, A., 1966. Wahrscheinlichkeitstheorie mit einem Anhang über Informationstheorie. VEB, Leipzig.
Ricotta, C., 2003. On parametric evenness measures. Journal of Theoretical Biology 222, 189–197.
Roberts, A., Varberg, D., 1973. Convex Functions. Academic Press, New York.
Schur, I., 1923. Über eine Klasse von Mittelbildungen mit Anwendung auf die Determinantentheorie. Sitzungsberichte der Berliner Mathematischen Gesellschaft 22, 9–20.
Shannon, C., 1948. A mathematical theory of communication. Bell Systems 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–additive measures of relative information. Journal of Combinatorics Information & System Sciences 2, 122–132.
Sharma, B., Taneja, I., 1975. Entropy of type (α, β) and other generalized measures in information theory. Metrika 22, 205–215.
Shorrocks, A., 1980. The class of additively decomposable inequality measures. Econometrica 48, 613–625.
Theil, H., 1967. Economics and Information Theory. North-Holland, Amsterdam.
Tilman, D., 1996. Biodiversity: Population versus ecosystem stability. Ecology 77, 97–106.
Tóthmérész, B., 1998. On the characterization of scale-dependent diversity. Abstracta Botanica 22, 149–156.

Tsallis, C., 1988. Possible generalization of Boltzmann–Gibbs statistics. Journal of Statistical Physics 52, 479–487.

Tsallis, C., 2004. What should a statistical mechanics satisfy to reflect nature. Physica D 193, 3–34.

Wang, G., Jiang, M., 2005. Axiomatic characterization of nonlinear homomorphic means. Mathematical Analysis and Applications 303, 350–363.

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

Woerheide, W., Persson, D., 1993. An index of portfolio diversification. Financial Services Review 2, 73–85.