How to measure own-group preference? A novel approach to a
sociometric problem

Non Peer-reviewed author version

How to measure own-group preference?
A novel approach to a sociometric problem

Leo Egghe
LUC, Universitaire Campus, B-3590 Diepenbeek, Belgium
& UA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium
E-mail: leo.egghe@luc.ac.be

and

Ronald Rousseau
KHBO, IWT, Zeedijk 101, B-8400 Oostende, Belgium
& UA, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium
E-mail: ronald.rousseau@khbo.be

Abstract

In this article we present a precise definition of the notion 'own-group preference" and characterize all functions capable of correctly measuring it. Examples of such functions are provided. The weighted Lorenz curve and the theory developed for it will be our main tools for reaching this goal. We further correct our earlier articles on this subject.

In the context of own-language preference, Bookstein and Yitzhaki proposed the logarithm of the odds-ratio as an acceptable measure of own-group preference. We now present a general framework within which the concept of own-group preference, and its opposite, namely own-group aversion, can be precisely pinpointed. This framework is derived form inequality theory and is based on the use of the weighted Lorenz curve.

The concept of own-group preference is an interesting notion with applications in different fields such as sociology, political sciences, economics,

management science and of course, the information sciences. Some examples are provided.

Introduction

The traveler immediately notices: there are many French cars on French roads, many German cars on German roads, and many American cars on American roads. Obviously, many people prefer (or are led to) buying a car manufactured by a company owned by businessmen of their own country. Similarly, readers prefer reading books written by novelists of their own country or region, and certainly published in their own language (translated or not). It even happens that scientists prefer reading articles in their own language, often published by fellow countrymen. Similarly, scientific authors publishing in a particular journal have a preference of citing other articles published in that same journal. This leads to the observation that journal self-citing rates are often very high (Fassoulaki et al., 2000; Nisonger, 2000; Ren & Rousseau, 2002).

In this article we present a precise definition of the notion 'own-group preference" and characterize all functions acceptable for its measurement. The weighted Lorenz curve and the theory developed for it will be our main tools for reaching this goal.

We studied own-group preference in earlier publications (Egghe et al, 1999; Egghe & Rousseau, 2000), but realized recently that we missed the point, and instead of studying preferences we have actually studied conditional relative use. This too will be explained further on in this article.

Is preference a notion characterized by absolute or by relative numbers?

It is obvious that preference is a relative notion. Saying that there are 800,000 French cars in France (a fictitious number) does not tell anything about the preference of the French to buy French. Similarly, stating that volume 53 of journal J contains 10231 references to articles published in journal J does not tell anything about J-writers' preference to cite their own journal.

If, moreover, during a certain period all readers borrow (for whatever reason) the double amount of books, this may not influence an own-group preference measure. We conclude that own-group preference must be expressed using relative, not absolute numbers. The term 'preference' clearly refers to an intrinsically relative notion.

A first step: percentage of own items used

The percentage of own items used clearly is a quantity that contains information about the own-group preference. In the context of the citation of articles written in the same language as that of the author, the term 'language self-reference' has been used (Yitzhaki, 1997). In the context of a journal J citing articles published in J, terms such as 'journal self-reference' or 'journal self-citation' are in use.

Although these percentages tell us something about relative use they are not really informative when it comes to (relative) preference. Indeed, in order to measure preference also the supply must be taken into account. This is the complete universe of available items, and the percentage of 'own' items

therein. If journal J publishes twice the amount of articles than journal T, then it is expected that journal J cites twice the amount of J-articles, than journal T does T-articles. Note also that it is of the utmost importance to define precisely the universe within which one works. In determining the own-language preference of the main public library of Antwerp or Amsterdam one may consider as universe the total amount of books available in this library, and the percentage of Dutch books among these; or one may consider the percentage of Dutch language books in the total world book production. These are two totally different objects of study!

Intuitively desirable properties for own-group preference

We will denote the fraction of *own items* (in the universe under consideration) by the symbol α and the fraction of *own items in the set of used items* by the symbol c. For the moment we will assume that $\alpha \leq c$. Indeed, if $\alpha > c$ then, obviously, there is no own-group preference. Hence we will work within the set $\{(\alpha,c) \in\ ]0,1] \times\ ]0,1]; \alpha \leq c\ \}$. This set will be called the own-group preference set. An element (α,c) of this set will be referred to as an own-group preference situation (or preference situation, for short). The set itself will be denoted as OGPS (own-group preference set). Note that if $\alpha = 0$ this would mean that the corresponding group does not exist (or at least has no elements). It is pointless to study such a group. Finally, we remark that if $\alpha = 1$, c is automatically also equal to 1.

When studying own-group preference it is (intuitively) natural to require that if α is kept fixed, then the higher c, the higher the own-group preference.

Similarly, when keeping c fixed, the higher α (but $\alpha \leq c$), the lower the own-group preference. Finally, when $\alpha = c$ we want the preference to be zero, since one uses the own group as expected by its own size. Putting the preference equal to zero in this situation is then a natural normalization requirement.

Finally, we do not want that a small change in one parameter results in a sudden disproportional change (a jump) in preference. In other words, we want an own-group preference measure to be continuous.

The weighted Lorenz curve solves the problem

In the case one is interested in how different concentration is with respect to a standard, one uses a weighted Lorenz curve. This standard can be an internal or an external standard (for details we refer to Rousseau (2001)). We briefly recall here the construction of such a weighted Lorenz curve (Theil, 1967; Patil & Taillie, 1982; Egghe & Rousseau, 2001). Let $S = (s_1, s_2, ..., s_N)$ denote the standard vector and let $X = (x_1, x_2, ..., x_N)$ denote the distribution vector that we want to compare with this standard. Note that indices must correspond. If, e.g., in a study of relative publication activity per country, X denotes numbers of publications and S denotes population then x_i and s_i must refer to the same country C_i. We assume, moreover, that none of the components of S is zero. In order to construct the Lorenz curve for comparison with a standard the components of both vectors are ranked in such a way that

$$\frac{x_1}{s_1} \geq \frac{x_2}{s_2} \geq ... \geq \frac{x_N}{s_N} \qquad (1)$$

Next we normalize the vectors X and S, leading to vectors C and W, where

$$c_i = \frac{x_i}{\sum\limits_{j=1}^{N} x_j} \quad and \quad w_i = \frac{s_i}{\sum\limits_{j=1}^{N} s_j} \tag{2}$$

Note that normalizing does not change the order. Finally, the weighted Lorenz curve is defined as the broken line connecting the origin (0,0) to the points with components

$$\left(\sum\limits_{j=1}^{i} w_j \ , \ \sum\limits_{j=1}^{i} c_j \right)_{i=1,...,N} \tag{3}$$

By construction this curve is always concave.

We will now apply this construction to the simple case where W = (α, 1- α) and C = (c, 1-c). Note that, because of the assumption that α ≤ c, c/α ≥ (1- c)/(1- α), as required by (1). Connecting the origin (0,0) with the point (α,c) and then this point with (1,1) yields the weighted Lorenz curve for this situation. Such a Lorenz curve will be denoted as L(α,c). Such curves introduce a partial order in the OGPS. If the weighted Lorenz curve associated with (α_1, c_1) is never situated strictly under the weighted Lorenz curve associated with (α_2, c_2) (and they do not coincide) then the own-group preference of (α_1, c_1) is intrinsically larger than that of (α_2, c_2). This is symbolically denoted as $(\alpha_1, c_1) \succ (\alpha_2, c_2)$. Intrinsic preference clearly enjoys the "intuitively desirable properties" mentioned above. Formally, this leads to the following definition.

Definition 1: own-group preference

$$L\left(\alpha_1, c_1\right) > L\left(\alpha_2, c_2\right) \Leftrightarrow \left(\alpha_1, c_1\right) \succ \left(\alpha_2, c_2\right)$$

Recall from the theory of (weighted) Lorenz curves that $L(\alpha_1,c_1) > L(\alpha_2,c_2)$ if and only if $L(\alpha_1,c_1)$ is at no point situated strictly under $L(\alpha_2,c_2)$, and there is at least one point (and hence infinitely many) where $L(\alpha_1,c_1)$ is strictly larger than $L(\alpha_2,c_2)$. This is illustrated in Fig.1. If weighted Lorenz curves intersect then the corresponding preference situations are intrinsically incomparable. From now on these weighted Lorenz curves will be called *own-group preference curves*.
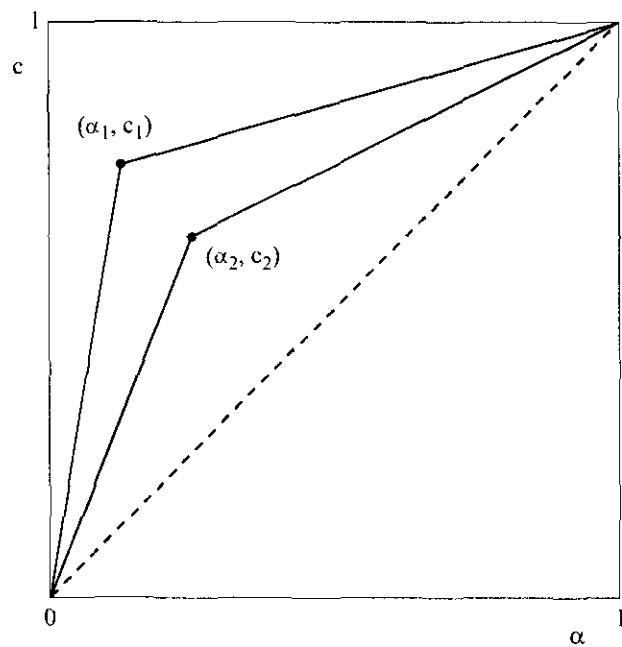

Fig. 1 Two own-group preference curves: $L(\alpha_1,c_1) > L(\alpha_2,c_2)$ hence $(\alpha_1,c_1) \succ (\alpha_2,c_2)$.

Following the theory of weighted Lorenz curves and the corresponding inequality measures we define an acceptable own-group preference measure as follows.

Definition 2: acceptable own-group preference measures

A function P, defined on OGPS, is an acceptable own-group preference measure if and only if it is a strictly increasing function for the partial order $\succ$. This may also be expressed by the phrase 'respecting the partial order determined by own-group preference curves'.

An equivalent mathematical expression for the partial order induced by own-group preference curves

Theorem 1

$(\alpha_1, c_1) \succ (\alpha_2, c_2)$ if and only if

$$\frac{c_1}{\alpha_1} \geq \frac{c_2}{\alpha_2} \text{ and } \frac{1-c_1}{1-\alpha_1} \leq \frac{1-c_2}{1-\alpha_2} \tag{4}$$

with at least one strict inequality.

Proof. Considering the corresponding own-group preference curves one can easily see that Theorem 1 is true. Indeed: by Figure 2, we see that $(\alpha_1, c_1) \succ (\alpha_2, c_2)$ if and only if $\beta_1 \geq \beta_2$ and $\gamma_1 \geq \gamma_2$ (with at least one inequality). This is equivalent with $\tan(\beta_1) \geq \tan(\beta_2)$ and $\tan(\pi/2 - \gamma_1) \leq \tan(\pi/2 - \gamma_2)$, or

$$\frac{c_1}{\alpha_1} \geq \frac{c_2}{\alpha_2} \text{ and } \frac{1-c_1}{1-\alpha_1} \leq \frac{1-c_2}{1-\alpha_2} \tag{4}$$
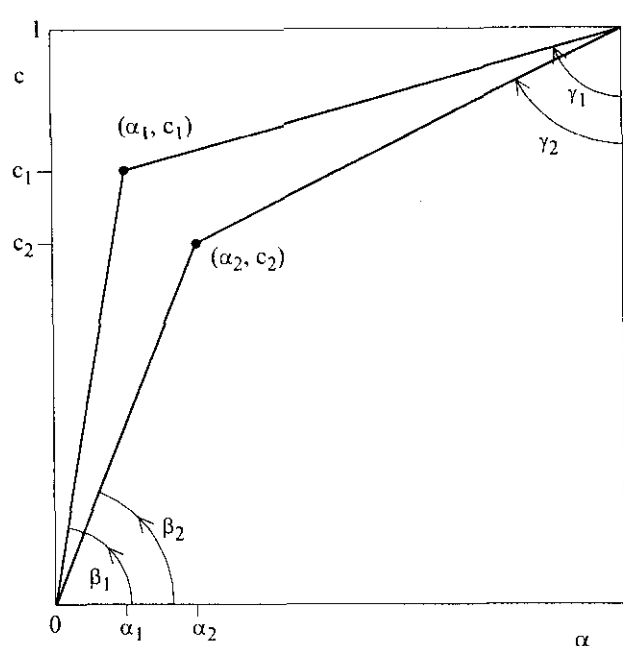
with at least one strict inequality.

Fig. 2 The own-group preference partial order can be expressed using the

angles β_j and γ_j , j = 1,2.

Using determinants, the double inequality (4) can be rewritten as follows:

$$\begin{vmatrix} \alpha_1 & c_1 \\ \alpha_2 & c_2 \end{vmatrix} \leq 0 \leq \begin{vmatrix} 1-\alpha_1 & 1-c_1 \\ 1-\alpha_2 & 1-c_2 \end{vmatrix} \tag{5}$$

Moreover, preference equality only occurs when the two inequality signs are actually equalities, or stated otherwise: there is a difference in own-group preference if at least one of the inequalities is strict.

Theorem 1 leads to the following important observation. The basic variables for own-group preference are not α and c, but s = c/α and t = (1-c)/(1-α). Clearly s $\in [1,+\infty[$ and t $\in]0,1]$. Before continuing our study of the own-group

preference, we will explicitate the one-to-one relation between the variables (α,c) and (s,t).

Lemma. The variables (α,c) can be expressed as a function of the variables (s,t) and vice versa. These relations are:

$$s = \frac{c}{\alpha} \quad \text{and} \quad t = \frac{1-c}{1-\alpha} \tag{6}$$

while conversely

$$\alpha = \frac{1-t}{s-t} \quad \text{and} \quad c = s\frac{1-t}{s-t} \tag{7}$$

Recall that always α < c.

Proof

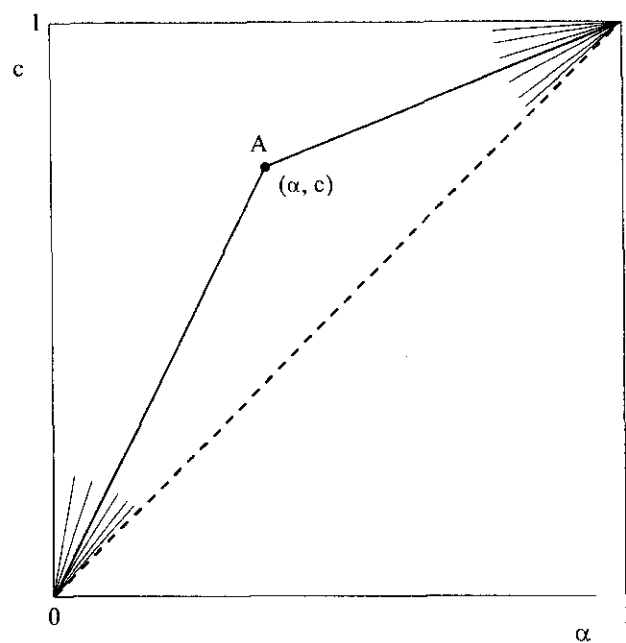The proof is just simple algebra and follows immediately from Fig.3. Hence it is omitted.

Fig.3 The point A determined by its coordinates (α,c) can also be determined by O and the slope of the line OA (i.e. s), and by I and the slope of the line AI (i.e. t).

Note that if $\alpha = c$ then $s = t = 1$. Conversely, however, if $s = 1 = t$ then the corresponding couple $(\alpha,c) = (\alpha,\alpha) = (c,c)$ is indeterminate.

Using the transformation formulae (6) and (7) any acceptable measure for own-group preference can equally well be expressed in terms of s and t, as in terms of α and c. Yet, a characterization of such measures is best expressed in terms of s and t. This is shown in the following theorem. Note also that the normalization requirement for a preference measure P (expressed as a function of s and t) now becomes $P(1,1) = 0$.

Theorem 2

A function $P(s,t)$, such that $P(1,1) = 0$ is an acceptable measure for own-group preference if and only if it is strictly increasing in s, and strictly decreasing in t.

Proof

We will first show that if P is an acceptable measure for own-group preference, it is strictly increasing in s and strictly decreasing in t. Assume first that t is fixed. If $s_1 > s_2$ then (s_1,t) corresponds to the point A in Fig.4, while (s_2,t) corresponds to point B. Clearly the preference curve OAI is situated strictly above the preference curve OBI. Consequently, $P(s_1,t) > P(s_2,t)$, showing that P is strictly increasing in s. (Note that this reasoning is also correct for t=0, and that if t=1 it is impossible to have $s_1 > s_2$.)
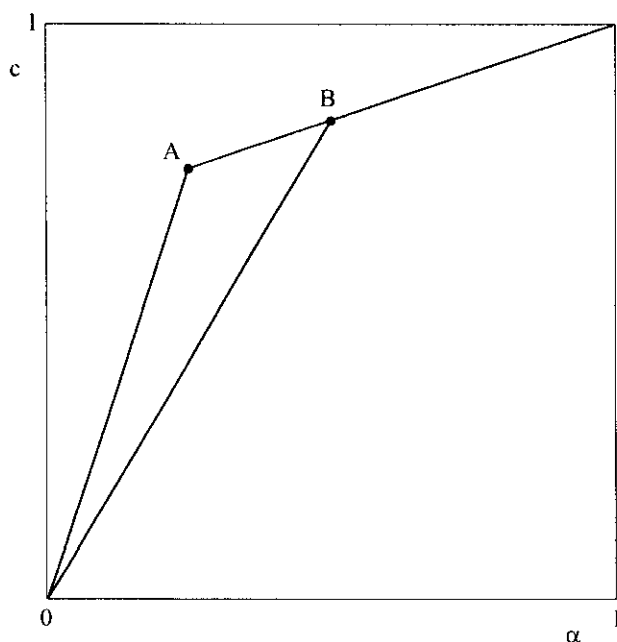

Fig.4 Two own-group preference curves with same t-value, but different s-values

Assume now that s is fixed. If $t_1 > t_2$ then (s,t_1) corresponds to the point C in Fig.5, while (s,t_2), corresponds to the point D. Clearly the preference curve OCI is situated strictly under the preference curve ODI. Consequently, $P(s,t_1)$ < $P(s,t_2)$ showing that P is strictly decreasing in t. (If s=1 it is impossible to have $t_1 > t_2$)

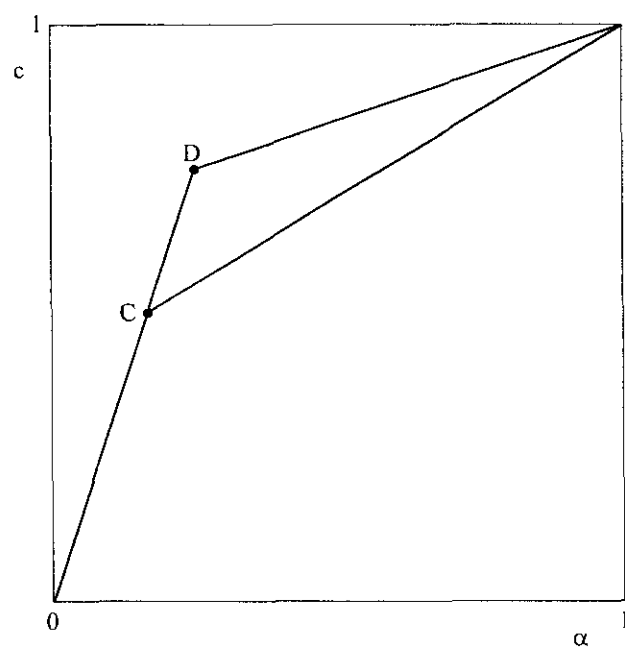Fig.5 Two own-group preference curves with same s-value, but different t-values

Finally, we have to show that if P is strictly increasing in s and strictly decreasing in t, then P is an acceptable own-group preference measure. This, however, follows immediately from Theorem 1 and the definition of an acceptable measure of own-group preference.

This proves the characterization theorem of acceptable measures.

Through the one-one relation between the variables (s,t) and (α,c) any acceptable measure P(s,t) can also be expressed as a function P(α,c) and vice versa. We will show next that such a function P(α,c) is increasing in α and decreasing in c. This agrees with our intuitive feeling. Yet, as we will show too, a function of two variables which is increasing in α and decreasing in t, is not necessarily an acceptable own-preference measure (according to our definition).

Theorem 3a. A necessary condition for P(s,t) to be an acceptable own-group preference measure, is that P(α,c) is strictly increasing in c, and strictly decreasing in α.

Proof.

Assume that P(s,t) is an acceptable own-group preference measure. Then we find, for P considered as a function of α and c:

$$\frac{\partial P(\alpha,c)}{\partial \alpha} = \frac{\partial P}{\partial s}\frac{\partial s}{\partial \alpha} + \frac{\partial P}{\partial t}\frac{\partial t}{\partial \alpha}$$

$$= -\frac{c}{\alpha^2}\frac{\partial P}{\partial s} + \frac{(1-c)}{(1-\alpha)^2}\frac{\partial P}{\partial t}$$

As $0 < c \le 1$, $\dfrac{\partial P}{\partial s}$ is strictly increasing and $\dfrac{\partial P}{\partial t}$ is strictly decreasing this expression is always negative, showing that P is strictly decreasing in α. Similarly,

$$\frac{\partial P(\alpha,c)}{\partial c} = \frac{\partial P}{\partial s}\frac{\partial s}{\partial c} + \frac{\partial P}{\partial t}\frac{\partial t}{\partial c}$$

$$= \frac{1}{\alpha}\frac{\partial P}{\partial s} - \frac{1}{1-\alpha}\frac{\partial P}{\partial t}$$

As $0 < \alpha < 1$, $\dfrac{\partial P}{\partial s}$ is strictly increasing and $\dfrac{\partial P}{\partial t}$ is strictly decreasing it follows

that P is strictly increasing in c.

Theorem 3b. This condition, while necessary, is not sufficient.

A simple counterexample for the opposite of Theorem 3a is given by the function $\pi_1(\alpha,c) = c/\alpha$. This function is strictly increasing in c, and strictly decreasing in α but yields the same values for the preference situations determined by C and D (Fig.5), although the preference curve determined by C is strictly larger than that determined by D.

It is even possible to give a stronger example. Consider the function $\pi_2(\alpha,c) = c/\alpha^2$. The function $\pi_2(\alpha,c)$ is strictly increasing in c and strictly decreasing in α. Expressed as a function of s and t, π_2 becomes (by equation (7)):

$$\pi_2(s,t) = \frac{s(s-t)}{1-t}$$

As $\dfrac{\partial \pi_2(s,t)}{\partial t} = \dfrac{s(s-1)}{(1-t)^2} \geq 0$, we see that π_2 is increasing in t, while an acceptable own-group preference measure must be strictly decreasing in t.

Examples of own-group preference measures

Any function of two variables, which is increasing in the first and decreasing in the second variable, and satisfied the normalization requirement that for s=t=1

its value is zero, may be used as an own-group preference measure. Simple examples are:

$$P_1(s,t) = (s-t)/t \qquad (8)$$

$$P_2(s,t) = s-t \qquad (9)$$

$$P_3(s,t) = \sqrt{s} \ln\left(\frac{1}{t}\right) \qquad (10)$$

And

$$P_4(s,t) = \sqrt{s} - \sqrt{t} \qquad (11)$$

We further note that any function respecting the general partial order induced by weighted Lorenz curves (the so-called measures of asymmetric relative concentration (Egghe, 2002; Egghe & Rousseau, 2001, Rousseau, 2001) can be adapted to yield an acceptable own-group preference measure. Here are some examples (using the general notation introduced above).

1) The asymmetric (or weighted) Theil measure:

$$Th_w(X) = \sum_{i=1}^{N} c_i \; \ln\left(\frac{c_i}{w_i}\right) \qquad (12)$$

This becomes: $Th_p(\alpha,c) = c \ln\left(\frac{c}{\alpha}\right) + (1-c)\ln\left(\frac{1-c}{1-\alpha}\right)$ $\qquad (13)$

As a function of s and t this is:

$$Th_p(s,t) = s\frac{1-t}{s-t}\ln(s) + t\frac{s-1}{s-t}\ln(t) \qquad (13')$$

Here we just put $Th_p(1,1) = 0$.

2) The asymmetric (or weighted) squared coefficient of variation:

$$V_w^2(X) = \sum_{i=1}^{N} \frac{(c_i - w_i)^2}{w_i} \tag{14}$$

This becomes:

$$V_p^2(\alpha,c) = \frac{(c-\alpha)^2}{\alpha(1-\alpha)} \tag{15}$$

As a function of s and t this is:

$$V_p^2 = (s-1)(1-t) \tag{15'}$$

3) The weighted Gini index:

$$G_w(X) = \frac{1}{2} \sum_{i=1}^{N}\sum_{j=1}^{N} \left| w_i c_j - w_j c_i \right| \tag{16}$$

This becomes $G_p(\alpha,c) = c - \alpha$, $\tag{17}$

As a function of s and t this is:

$$G_p(s,t) = \frac{(1-t)(s-1)}{s-t} \tag{17'}$$

Also here we put $G_p(1,1) = 0$.

The interpretation of this index is similar to that of the usual Gini index, namely twice the area between the own-group preference curve and the diagonal.

4) The normalized length of the own-group preference curve:

$$LOR_w(X) = \sum_{i=1}^{N}\sqrt{c_i^2 + w_i^2} - \sqrt{2} \tag{18}$$

and this finally becomes

$$LOR_p(\alpha,c) = \sqrt{\alpha^2 + c^2} + \sqrt{(1-\alpha)^2 + (1-c)^2} - \sqrt{2} \tag{19}$$

As a function of s and t this is:

$$LOR_p(s,t) = \frac{1-t}{s-t}\sqrt{1+s^2} + \frac{s-1}{s-t}\sqrt{1+t^2} - \sqrt{2} \tag{19'}$$

Applying any strictly increasing function mapping zero to zero to these measures is again an acceptable measure for own-group preference. The Gini preference, for instance, leads to $(c-\alpha)^2$ or $\sqrt{c-\alpha}$ or, in general, $(c-\alpha)^\beta$, with $\beta > 0$. These are all acceptable own-group preference measures. Expressed as functions of s and t this leads to functions of the form

$$\left(\frac{(1-t)(s-1)}{s-t}\right)^\beta.$$

The odds ratio proposed by Bookstein and Yitzhaki (1999a,b) is a good own-group preference measure, which can easily be transformed to satisfy the normalization requirement that we have imposed. Note that in previous articles we rejected the odds ratio. The reason for this is that we were actually studying a slightly different notion (see further). The odds ratio is defined as:

$$\frac{c(1-\alpha)}{(1-c)\alpha} \tag{20}$$

It takes values between 0 and 1 when $c < \alpha$, between 1 and infinity, when $c > \alpha$ and its value is 1 for $\alpha = c$. Hence subtracting the value one or taking logarithms leads to acceptable own-group preference measures:

$$O_1(\alpha,c) = \frac{c(1-\alpha)}{\alpha(1-c)} - 1 = \frac{c-\alpha}{\alpha(1-c)} \tag{21}$$

and

$$O_2(\alpha,c) = \ln\left(\frac{c(1-\alpha)}{\alpha(1-c)}\right) \tag{22}$$

We note that Bookstein and Yitzhaki (1999a,b) already suggested such normalizations. Expressed as functions of s and t these measures become:

$$O_1(s,t) = (s-t)/t \quad \text{and} \quad O_2 = \ln(s/t) \tag{23}$$

Note that O_1 is nothing but the measure P_1 introduced above as equation (8).

Taking an arithmetic variant of (20) leads to another acceptable measure:

$$A = \frac{c}{\alpha} - \frac{1-c}{1-\alpha} = \frac{c-\alpha}{\alpha(1-\alpha)} \tag{24}$$

Expressed as a function of s and t this becomes: A(s,t) = s − t, or the function P_2 introduced in equation (9).

Note that only the Gini preference and the normalized length of the preference curve are bounded measures (by 1 and 2-√2, respectively). Dividing equation (19) or (19') by 2-√2 leads to a measure bounded by 1.

Own-group aversion

What happens if c < α? In this case the own-group is used less than expected. One may say that there is an own-group aversion. Note that we will use this terminology as a purely technical term. No negative connotation should be attached to it.

One may draw a curve in the same way as in the preference case. Instead of a concave curve, we now have a convex curve. See Fig.6.
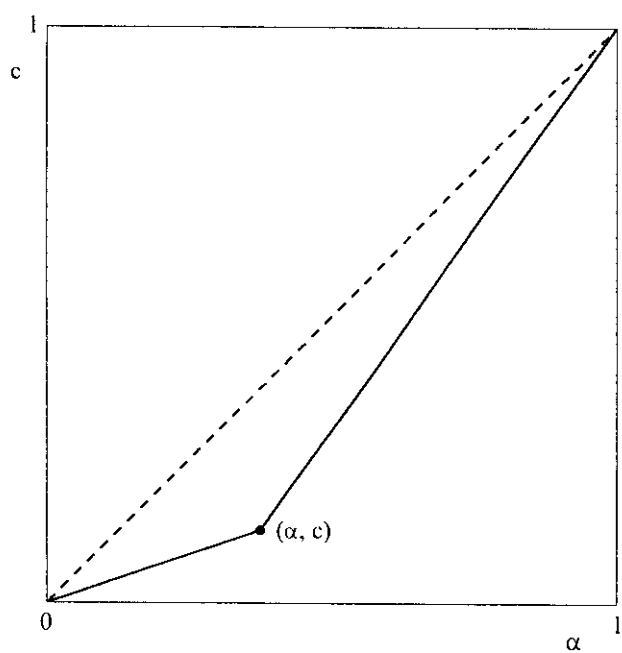
Fig.6 An own-group aversion curve

This curve corresponds to a weighted Lorenz curve, where now (using again the general notation):

$$\frac{x_1}{s_1} \leq \frac{x_2}{s_2} \leq \ ... \ \leq \frac{x_N}{s_N} \tag{25}$$

It is well-known that this representation is equivalent with the one where these fractions are ranked from largest to smallest. Consequently, own-group aversion can be measured by the same functions as own-group preference (or their absolute values if necessary). The Gini own-group aversion measure, for instance, is simply

$$G_A(\alpha,c) = \alpha - c > 0 \tag{26}$$

Clearly $G_A = |G_P|$.

Further $Th_A = Th_P$, $V_A^2 = V_P^2$, $LOR_A = LOR_P$, and the aversion analogues derived from the odds ratio are: $O_{1A} = \dfrac{\alpha - c}{\alpha(1-c)}$ and $O_{2A} = \ln\left(\dfrac{\alpha(1-c)}{c(1-\alpha)}\right)$.

Note that functions such as Th_A and Th_P are described by the same expression, but that they have a different domain. So, strictly speaking they are different functions.

Sensitivity

In order to choose among the many preference (or aversion) measures one may consider additional requirements. If P(s,t) denotes a general acceptable own-group preference measure, then a possible sensitivity requirement is that P changes faster for small s than for larger ones (with t fixed). Mathematically this becomes:

$$\forall t \in]0,1]: \left|\frac{\partial P(s,t)}{\partial s}\right| = \frac{\partial P(s,t)}{\partial s} \text{ must be strictly decreasing in s.} \tag{27}$$

$$\text{or: } \frac{\partial^2 P(s,t)}{\partial s^2} < 0$$

This condition can be considered as a kind of law of diminishing returns. A similar requirement can be imposed on the derivative with respect to t.

$$\forall s \in]0,1]: \left|\frac{\partial P(s,t)}{\partial t}\right| = -\frac{\partial P(s,t)}{\partial t} \quad \text{must be strictly decreasing in t.}$$

$$\text{or: } \frac{\partial^2 P(s,t)}{\partial t^2} > 0 \tag{28}$$

In case one is interested in such a measure then the logarithm of the odds-ratio (23) is an acceptable measure satisfying these two extra requirements. $P_3(s,t) = \sqrt{s}\ln\left(\frac{1}{t}\right)$ and $P_4(s,t) = \sqrt{s} - \sqrt{t}$ too satisfy these sensibility requirements.

What did we study in previous articles?

In (Egghe et al, 1999; Egghe & Rousseau, 2000), we studied what we called the 'relative own-language preference', following Yitzhaki's terminology (Yitzhaki, 1997). Now we realize that this is a wrong terminology. In those articles we considered the following partial order (denoted as -< , and where the symbols α and c have the same meaning as in this article):

$$(\alpha_1, c_1) \prec (\alpha_2, c_2) \quad \text{if and only if}$$

$$(\alpha_2 \leq \alpha_1 \text{ and } c_1 \leq c_2) \text{ or } c_1 = 0 \text{ or } (\alpha_1, c_1) = (1,1)$$

Note also that in our earlier articles there was no restriction such as $\alpha \leq c$ either.

Denoting $1-\alpha$ by δ (the fraction of 'foreign' items), this means:

$$(\alpha_1, c_1) \prec (\alpha_2, c_2) \text{ if and only if}$$

$$(\delta_1 \leq \delta_2 \text{ and } c_1 \leq c_2) \text{ or } c_1 = 0 \text{ or } \delta_1 = 0$$

This notion does not express preference, but rather relative use given a choice to make use of foreign items. Indeed, when there is no use ($c=0$) or no choice possible ($\delta = 0$) then this notion is at its minimum. In other cases the notion increases with larger use and more options to chose outside the own group (own language). With this new interpretation nothing in our previous articles is logically or structurally wrong, hence – mathematically - the results still hold, albeit with another interpretation.

Examples

A new look at own-language preference.

In a comparative study of language preference in the field of sociology, Yitzhaki (1997) examined the language of references cited in regular original research articles published between 1989 and 1994 in nine different-nationality sociology journals: American, British, German and French (see Table 1). The α-value was calculated on the basis of two alternative estimates. Estimate 1 was based on a statistical analysis of the languages of the publications included in **SOCIOFILE** CD-ROM 1/74 - 12/95, while estimate 2 was based on the assumption that only 50% of the non-English items were included in that database.

Table 1 Nine sociology journals (as studied in previous articles) and their abbreviations

Theory, Culture and Society (UK)	TCS
American Sociological Review (USA)	ASR
Sociology (UK)	SOC
Sociological Review (UK)	SR
British Journal of Sociology (UK)	BJS
Revue Française de Sociologie (F)	RFS
L' Année Sociologique (F)	AS
Kölner Zeitschrift für Soziologie und Sozial Psychologie (G)	KZS
Sozialwissenschaftliche Information für Unterricht und Studium (G)	SI

Table 2. Own-preference values for the nine sociology journals presented in

Table 1

	alpha1	c	Th		Var-sq		Gini	
TCS	0.814	0.896	0.026	9	0.044	9	0.082	9
ASR	0.814	0.986	0.153	8	0.195	8	0.172	8
SOC	0.814	0.988	0.159	7	0.200	7	0.174	7
SR	0.814	0.993	0.174	6	0.212	6	0.179	6
BJS	0.814	0.995	0.182	5	0.216	5	0.181	5
RFS	0.048	0.645	1.326	4	7.800	4	0.597	4
ASR	0.048	0.675	1.435	3	8.603	3	0.627	2
KZS	0.03	0.633	1.573	2	12.495	2	0.603	3
SI	0.03	0.855	2.589	1	23.389	1	0.825	1

			LOR		O1		A	
TCS			0.009	9	0.969	9	0.542	9
ASR			0.051	8	15.093	8	1.136	8
SOC			0.052	7	17.813	7	1.149	7
SR			0.056	6	31.415	6	1.182	6
BJS			0.057	5	44.472	3	1.195	5
RFS			0.249	4	35.035	5	13.065	4
ASR			0.268	2	40.192	4	13.721	3
KZS			0.257	3	54.768	2	20.722	2
SI			0.422	1	189.655	1	28.351	1

	alpha2	c	Th		Var-sq		Gini	
TCS	0.687	0.896	0.123	9	0.203	9	0.209	9
ASR	0.687	0.986	0.313	6	0.416	7	0.299	7
SOC	0.687	0.988	0.320	5	0.421	6	0.301	6
SR	0.687	0.993	0.339	4	0.435	5	0.306	5
BJS	0.687	0.995	0.348	3	0.441	4	0.308	4
RFS	0.355	0.645	0.173	8	0.367	8	0.290	8
ASR	0.355	0.675	0.211	7	0.447	3	0.320	3
KZS	0.05	0.633	1.258	2	7.156	2	0.583	2

SI	0.05	0.855	2.155	1	13.643	1	0.805	1

	LOR		O1		A	
TCS	0.045	9	2.925	7	0.972	9
ASR	0.101	6	31.088	6	1.390	7
SOC	0.102	5	36.511	4	1.400	5
SR	0.106	4	63.631	3	1.423	4
BJS	0.108	3	89.665	2	1.432	3
RFS	0.058	8	2.301	9	1.267	8
ASR	0.071	7	2.774	8	1.398	6
KZS	0.239	2	31.771	5	12.274	2
SI	0.403	1	111.034	1	16.947	1

The two O-measures always yield the same ranking, so only one of them is shown. This follows from the fact that they are just monotone transformations from the basic odds ratio. It seems though that they are too sensitive for small changes. Indeed, they are the only ones that do not place KZS at the second place, but fifth instead. This seems counterintuitive as with an alpha-value of 5% and a relative use value, c, of 63.3% it is obvious that this journal has a high own-language preference, deserving a second place in the ranking. For the first publication share estimates the British *Theory, Culture and Society* has the lowest own-language preference, followed by the other English-language journals. The German journals have the highest own-language preference, while the French journals form the middle group. For the second estimates, however, we have a different situation. *Theory, Culture and Society* still has the lowest own-language preference, and the German journals have the highest (except for the measures based on the odds-ratio). Rankings for the other journals depend on the used measures, with the French *Revue Française de Sociologie* having the second lowest own-language preference.

Another example. In an investigation of book use at the library of R.R.'s (Dutch speaking) institute (Rousseau & Vandegehuchte, 1995) 1916 of the 5866 available books were in Dutch. Yet among the borrowed books 534 in a total of 1265 were Dutch. Hence α = 0.327 and c = 0.422. The corresponding Gini preference value is simply c - α or 0.095, a value which is much smaller than most values for own-language preference in sociological journals. It would be interesting to know if similar results hold for other libraries, in Flanders as well as abroad.

Journal self-citation (self-citing as well as self-cited data) too can be seen as a form of own-group preference. One may calculate, e.g. for a particular year, or a specific issue, the fraction of references to the journal, J, itself. Hence, the parameter c(J) is relatively easy to determine. In some cases it can even be obtained from ISI's *Journal Citation Reports*. It is much more difficult to determine a reasonable universe and calculate the fraction of articles published in journal J. In their study of semantically related journals Pudovkin and Garfield (2002) use the number of articles published in journal J (in one particular year). When comparing journals in the same pool (the same universe) this means that if the used own-group preference measure P(α,c) can be rewritten as

$$P(\alpha,c) = F(T)^*P(a,c) \qquad (29)$$

where F(T) denotes a function only depending on the total number of items in the universe (T), and α=a/T, then this might be an acceptable procedure.

Pudovkin and Garfield (2002) use P(α,c) = c/α (actually c/a) as a measure for journal self-relatedness. As explained above, we do not consider this as an

acceptable measure for own-group preference. We further note that none of the measures proposed by us is of the form suggested by equation (29) and that we started our investigation by stating that an own-group preference measure should be based on relative numbers, not on absolute ones. This means that it is probably necessary 1) to determine a universe explicitly, and 2) establish a reasonable estimate for the fraction of journal J-publications in this universe. We leave this as an open problem.

Conclusion

Yitzhaki (1997) determined one important component for own-group preference, in the case of language use. Then Bookstein and Yitzhaki (1999a,b) correctly proposed the odds-ratio as an acceptable measure of own-group preference. At the same time Egghe and Rousseau – with some help of Yitzhaki - studied a slightly different concept, namely relative use given a choice to make use of foreign items (Egghe et al., 1999). They however made an attempt to put such a study in a general framework, using partial order relations and a variation of the Lorenz curve concept (Egghe & Rousseau, 2000). We think we have now, finally, precisely pinpointed the concepts of own-group preference, and its opposite own-group aversion. They arise as special cases of a partial order derived from the general weighted Lorenz curve.

It is clear that the concept of own-group preference can be applied to many cases in different fields such as sociology, political sciences, economics, management science and of course, the information sciences. We express

the hope that this will be the case indeed, so that more experience will be gained in the use of these measures.

Acknowledgement. R.R. thanks Profs. Irene Wormell (Högskolan i Borås, Sweden) and Liang Liming (Henan Normal University, China) for their hospitality during his visits. It was while discussing own-language preference with them that he realized that something was wrong with the earlier approach.

References

BOOKSTEIN, A., M. YITZHAKI (1999a), Own-language preference: a new measure of 'relative language self-citation'. In: *Proceedings of the seventh conference of the International Society for Scientometrics and Informetrics*, Macias-Chapula (ed.), Colima, Universidad de Colima, pp. 65-74.

BOOKSTEIN, A., M. YITZHAKI (1999b), Own-language preference: a new measure of 'relative language self-citation', *Scientometrics*, 46, 337-348.

EGGHE,L. (2002), Construction of concentration measures for general Lorenz curves using Riemann-Stieltjes integrals, *Mathematical and Computer Modelling*, 35: 1149-1163.

EGGHE, L., R.ROUSSEAU (2000), Partial orders and measures for language preferences, *Journal of the American Society for Information Science*, 51: 1123-1130.

EGGHE,L., R.ROUSSEAU (2001), Symmetric and asymmetric theory of relative concentration and applications, *Scientometrics*, 52(2), 2001, 261-290.

EGGHE,L., R.ROUSSEAU, M.YITZHAKI (1999), The "own-language preference": measures of "relative language self-citation", *Scientometrics*, 45:217-232.

FASSOULAKI, A., A.PARASKEVA, K.PAPILAS, G.KARABINIS (2000), Self-citations in six anaesthesia journals and their significance in determining the impact factor, *British Journal of Anaesthesia*, 84:266-269.

NISONGER, T.E. (2000), Use of the Journal Citation Reports for serials management in research libraries: an investigation of the effect of self-citation on journal rankings in library and information science and genetics, *College & Research Libraries*, 60:263-275.

PATIL, G.P., C. TAILLIE (1982), Diversity as a concept and its measurement, *Journal of the American Statistical Association*, 77, 548-561.

PUDOVKIN, A.I., E. GARFIELD (2002), Algorithmic procedure for finding semantically related journals, *Journal of the American Society for Information Science and Technology*, 53, 1113-1119.

REN, S., R. ROUSSEAU (2002), International visibility of Chinese scientific journals, *Scientometrics*, 53: 389-405.

ROUSSEAU, R. (2001), Concentration and evenness measures as macro-level scientometric indicators (in Chinese). In: *Research and university evaluation* (Ke yan ping jia yu da xue ping jia), Jiang Guo-hua (ed.), 2001, Red Flag Publishing House, Beijing, p.72-89. [English version available from the author.]

ROUSSEAU,R., P.VANDEGEHUCHTE (1995), Books and their users: socio-cultural and linguistic aspects in an engineering school. *Library Science with a slant to documentation and information studies*, 32: 143-150.

THEIL, H. (1967), *Economics and information theory*. Amsterdam: North-Holland.

YITZHAKI,M. (1997), The language preference in sociology: measures of language self-citation, relative own-language preference indicator, and mutual use of languages. In: B.Peritz & L.Egghe (eds.): *Proceedings of the Sixth Conference of the International Society for Scientometrics and Informetrics*. Jerusalem, The Hebrew University, pp. 505-515.