

# UC Irvine

## UC Irvine Previously Published Works

### Title

Effective Size and Number of Components

### Permalink

<https://escholarship.org/uc/item/0hm6r8sk>

### Journal

Sociological Methods & Research, 10(1)

### ISSN

0049-1241

### Authors

Taagepera, R  
Grofman, B

### Publication Date

1981

### DOI

10.1177/004912418101000104

Peer reviewed

*Two families of indices are defined which may be used to characterize the effective number of components in systems with unequal component sizes. A general formula is stated relating size, effective size, and effective number of components. Special cases of this formula are considered which yield formulae identical or closely related to a variety of other expressions including entropy, the Greenberg-Liebers index of diversity (also known as the Rae-Taylor fractionalization index), the Herfindahl-Hirschman concentration index, the coefficient of variation, the ordinary mean, the weighted mean and the harmonic mean. Applications of these formulae are considered for a variety of problems, including measurement of population and GNP concentrations, distinguishing between density and crowding, and reconciling differing student and faculty perceptions of average class size.*

## Effective Size and Number of Components

REIN TAAGEPERA  
BERNARD GROFMAN  
*University of California—Irvine*

### WEIGHTED AND UNWEIGHTED AVERAGES

The purpose of this article is to demonstrate the basic equivalence of two very different-looking approaches to the problem of characterizing systems with unequal component sizes, and to discuss the theoretical and operational consequences of that equivalence.

Feld and Grofman (1977, 1980) have pointed out that the average class size experienced by students differs from that seen by the administration. For  $S$  students divided among  $N$  classes, with  $S_i$  students in the  $i^{\text{th}}$  class, the administration sees an average

---

AUTHORS' NOTE: *An earlier version of this paper was presented at the annual meeting of the Public Choice Society, Charleston, SC, March 17-19, 1979.*

class size given by: mean  $S_1 = S/N$ . But students see an average size given by: mean  $S_2 = \sum S_i^2/S$ , since a size of  $S_i$  is experienced by a fraction  $S_i/S$  of the students. As a simple extreme case, consider a school offering only 3 courses with nonoverlapping enrollments of 10, 10 and 100 students, respectively. The administration may see a quite reasonably small average class size of  $(10 + 10 + 100)/3 = 40$  students. Yet the students experience a drastically larger average class size of  $(10^2 + 10^2 + 100^2)/(10 + 10 + 100) = 85$  students, because nearly all students attend the crowded 100-student class. One can visualize a student-administration conflict developing, with the students complaining of a very real overcrowding that the administration honestly cannot see. Feld and Grofman further discussed various statistical aspects of the problem, connecting the effective grass-roots average to the standard deviation.<sup>1</sup>

In general, members of a system experience, on the average, larger components than are seen by outsiders, because larger components are experienced by more members. Thus the outsider may see in North America three countries (the United States, Mexico, and Canada) with an average population of  $S_1 = (220 + 70 + 25)/3 = 105$  million people, but the average experience of the insider is mean  $S_2 = (220^2 + 70^2 + 25^2)/315 = 171$  million, reflecting the fact that most North Americans experience the United States, rather than Mexico or Canada.

In some cases, such as the class size situation pointed out by Feld and Grofman (1977), it may be easy to get confused in the choice between mean  $S_1$  and mean  $S_2$ . In other cases, uncertainty about the number of components in the system may give us a preference for mean  $S_2$ , since mean  $S_2$  does not depend greatly upon  $N$ . For example, geographically, the Bahamas, Bermuda, and Greenland also belong to North America. We neglected them because they are easy to forget, being of negligible population size compared to the three major countries.

If we include the Bahamas (.2 million), Bermuda (.06 million) and Greenland (.06 million) into our calculations, our summit-level average drops dramatically to 53 million:

$$S_1 = (25 + 220 + 70 + 0.2 + 0.06 + 0.06)/6 = 53.$$

However, the grass-roots average remains around 171 million:

$$S_2 = (25^2 + 220^2 + 70^2 + .2^2 + .06^2 + .06^2)/(25 + 220 + 70 + 2 + .06 + .06) = 171.$$

This is a type of average that does not change, regardless of whether very small components are counted in or out. Operationally, this stability is a very useful characteristic because the number of small components in some system might not be known with certainty. Beyond this operationally desirable feature, it can also be argued that mean  $S_2$  is, indeed, the average size effectively seen by participants, as opposed to mean  $S_1$ , which may be the average seen by outsiders or supervisors. In other words, mean  $S_1$  is the arithmetical mean over all components of the system (countries, in the case above), while the "effective size" mean  $S_2$  can be said to be the arithmetical mean over all the basic building blocks out of which the components are made (people, in the case above).

The general formula for the effective grass-roots average can be expressed as

$$\bar{S}_2 = \frac{\sum_{i=1}^N S_i^2}{\sum_{i=1}^N S_i} = \frac{\sum_{i=1}^N S_i^2}{S}, \quad [1]$$

where  $S$  is the total size,  $S_i$  is the size of the  $i^{\text{th}}$  component, and  $N$  is the total number of components. Small components can be omitted (thus altering  $N$ ) without much affecting the outcome. In contrast, the usual arithmetical mean (the summit-level average) depends very much on the value of  $N$  chosen:

$$\bar{S}_1 = \sum_{i=1}^N S_i / N = S/N. \quad [2]$$

Let us proceed now to what may look like quite different problems—those of industrial concentration and political fractionalization.

*CONCENTRATION AND  
EFFECTIVE NUMBER OF COMPONENTS*

If  $P_i = S_i/S$  is the fractional share of the  $i^{\text{th}}$  component, then the degree of concentration of the system can be expressed by the well-known Herfindahl-Hirschman concentration index (Hirschman, 1945; modification used here is by Herfindahl, 1950):

$$HH = \sum_{i=1}^N P_i^2 = \sum_{i=1}^N (S_i/S)^2. \quad [3]$$

E.g., in 1976, General Motors produced 58% of all U.S. cars; Ford produced 24%; Chrysler, 16%; and American Motors, 2%. The concentration of the U.S. automobile industry would be  $HH = .58^2 + .24^2 + .16^2 + .02^2 + \dots = .34 + .06 + .03 + .004 + \dots = .42$ . (The ellipses indicate the visibly negligible contribution to the overall concentration by minor producers.) The HH index varies from practically zero (for extreme fractionalization) to one (for complete monopoly by a single component). Conversely, fractionalization could be expressed by  $1 - HH$ , an index known in sociology as diversity index A (Greenberg, 1956; Gibbs and Martin, 1962; Lieberman, 1969), and in political science as the fractionalization index (Rae and Taylor, 1970):

$$A = 1 - HH. \quad [4]$$

Relationships (or lack of relationship) between HH and an amazing jungle of other indices also used to express dispersion, entropy, inequality, segregation, and the like, have been discussed recently by many a scholar (Allison, 1978; Blau, 1977; Taagepera, 1980, to name but the most recent ones). A list of 20 indices and their links to HH is given by Taagepera and Ray (1977). They will not be rediscussed here, with the following exception.

It has been pointed out by Laakso and Taagepera (1978, 1979) that the reciprocal of HH has an intuitively appealing interpre-

tation: It can be construed as an "effective" number ( $N_2$ ) of components, i.e., the number of *equal* components which would yield the same concentration index as do the actual unequal components:<sup>2</sup>

$$N_2 = 1/HH = 1/(1-A) = 1/\Sigma P_i^2 = S^2/\Sigma S_i^2. \quad [5]$$

Thus the effective number of U.S. car manufacturers in 1976 would be  $N_2 = 1/HH = 1/.42 = 2.3$ , which means that the concentration (in terms of HH) is the same as it would be if there were two to three equal-size companies.<sup>3</sup>

As another example, consider an election where 8 parties receive, respectively, 5, 4, 2, 1, .5, .2, .1, and .05 million votes, a configuration which sometimes happens in countries with proportional representation rules. Clearly fewer than 8 parties really count in the political process, but it is hard to specify a cut-off point. The effective number formula yields  $N_2 = (5 + 4 + \dots + .05)^2 / (5^2 + 4^2 + \dots + .05^2) = 3.6$ , which suggests that the system in some of its aspects may operate in the same way as if it had 3 to 4 equal-size parties (provided, of course, that electoral rules, ideological affinities, and coalition ability do not alter the relative strengths at the parliament seats and government formation levels).<sup>4</sup> The effective number  $N_2$  is always smaller than the actual number of components (unless all components are of equal size, in which case the effective number is equal to the actual one).

#### *COMBINING THE LAAKSO-TAAGEPERA AND FELD-GROFMAN APPROACHES*

The problem areas addressed by the Feld-Grofman class-size paradox and by the Laakso-Taagepera effective size of political parties were quite different, and so were the initial notations and paths of reasoning. Thus, the resulting indicators looked quite different, serving different purposes. However, these two seemingly different approaches are theoretically intertwined. This becomes clear if we divide the total size  $S$  by the Laakso-Taage-

pera effective number of components ( $N_2$ ), in order to obtain the corresponding size of the effective components. The result is

$$S/N_2 = SHH_2HH = S \sum S_i^2 / S^2 = \sum S_i^2 / S = \text{mean } S_2. \quad [6]$$

Expression 6 is precisely the grass-roots effective size obtained by Feld and Grofman. Equivalently,

$$S = N_2(\text{mean } S_2). \quad [7]$$

Given any two of: the total size, the effective size, and the effective number of components, the third one can be calculated, and so can the Herfindahl-Hirschman concentration index  $HH$ , the Lieberman diversity (or Rae-Taylor fractionalization) and various other related indices. We thus have a wide intercalculability between indices based on widely different approaches. Each of these approaches is conceptually strengthened by the interconnection between them.

Now let us apply these notions to a variety of sociopolitical data. First, let us consider the effective size of French communes, which range in size from Paris down to mountain hamlets with a few dozen people. There are 38,000 communes for a total of about 50 million people, so that the average commune size is mean  $S_1 = 1300$  people. However, most French people live in communes much larger than this average, which depends heavily on the numerous small communes. Any nationwide program geared mainly for the "average" 1300-person settlement would miss the majority. The grass-roots effective size turns out to be mean  $S_2 = 165,000$ , the size of Grenoble, Toulon, or Montpellier. Taking into account the  $N_2 = 300$  largest French communes means including towns down to about 12,000 inhabitants. Concentration is  $HH = .0033$ . (The  $HH$  value of .033 listed in Taylor and Hudson [1972: 222] is off by one decimal place.)<sup>5</sup>

Next, let us consider the world population distribution by countries around 1975. The total population was close to 4000 million, and there were about 160 sovereign or semisovereign countries, leading to mean  $S_1 = 25$  million. Yet the 5 largest

countries (China, India, USSR, the United States, and Indonesia) accounted for more than half of the world population. It is found that the effective average size, i.e., the country size of the average person, is mean  $S_2 = 316$  million, corresponding to an effective number of  $N_2 = 12.7$  countries, and to a concentration of  $HH = .08$ .<sup>6</sup> If the effective size of 316 million people per country seems intuitively too large, the reason may be that we are discounting the enormous masses experiencing life in China or India, because of their low GNP. We may thus want to repeat the calculations with GNP instead of population. The world total GNP in 1975 was \$5989 billion, without China (CBS, 1978: 175). Assuming that China's per capita GNP was about the same as India's (\$158), the total world GNP comes to  $S = \$6123$  billion and  $S_1 = \$38$  billion. Country-by-country data yield mean  $S_2 = \$584$  billion (a GNP surpassed only by the United States and the USSR),  $N_2 = 10.5$ , and  $HH = .10$ . Thus, the world concentration of GNP is slightly higher than that of population. (We do not consider here the additional concentration of wealth among groups or individuals within each country, nor population density variations within a country, nor the degree of willingness of various national groups to belong to the country that includes them.) The countries heading the GNP and population lists are, of course, different. Moreover, the grass-roots visualization of mean  $S_2$  given at the start of this paper would not be suitable in the case of GNP: It would involve dollars (rather than people) reporting on other dollars they see in their country's GNP. Thus the concept of mean  $S_2$  (or of  $N_2$  and  $HH$ ) becomes more abstract without losing its basic usefulness.

The grass-roots approach still could be maintained by asking how many people see how large GNP's. This leads to a mixed expression in two variables, with GNP averaged over all people:

$$\overline{(\text{GNP})}_P = \Sigma (\text{GNP})_i P_i / \Sigma P_i = \$213 \text{ billion} \quad [8]$$

for the world GNP. This is the size of the national economic system in which the average person lives or, in statistical physics terminology (Kittel, 1958: 9), the ensemble average of GNP over



the world population. Using general formulae analogous to those in equation 8, we could again calculate an effective number of countries (in terms of economic power):

$$N_{\text{GNP(P)}} = (\text{GNP})/(\text{GNP})_P = 29,$$

and also a concentration:

$$\text{HH}_{\text{GNP(P)}} = 1/N = .035.$$

This approach to effective number of countries takes into account the disparities both in population size and in size of national economy. At this mixed level the world is much less concentrated than on the population or the GNP level separately. The reduced concentration expresses the fact that the high-population and the high-GNP countries are not the same.

### GENERALIZED FAMILIES OF INDICES

Both mean  $S_1$  and mean  $S_2$  are special cases of the generalized family of expressions of the type

$$\bar{S}_n = \frac{\sum_{i=1}^N (S_i)^n}{\sum_{i=1}^N (S_i)^{n-1}}. \quad [9]$$

Another member of this family (for  $n = 0$ ) is

$$S_0 = N / \sum_{i=1}^N (1/S_i), \quad [10]$$

i.e., the inverse of the arithmetic mean of the inverses of component sizes, which is known as the *harmonic mean*.<sup>7</sup>

One could easily take off from the example of equation 8 into a new round of generalization from one variable (as in equation 9) to many variables:

$$\bar{S}(n; T^m, U^p, \dots) = \frac{\Sigma (S_i)^n T_i^m U_i^p \dots}{\Sigma (S_i)^{n-1} T_i^m U_i^p \dots} \quad [11]$$

In this paper we will not consider the possible uses of combinations other than the aforementioned ones: mean  $S(1; T^0, U^0, \dots) = \text{mean } S_1$ , mean  $S(2; T^0, U^0, \dots) = \text{mean } S_2$ , and mean  $S(1; T^1, U^0, \dots)$ , the latter being the form of mean  $(\text{GNP})_p$  in 8 (using  $S = \text{GNP}$  and  $T = \text{population}$ ). We believe, however, that such indices may be useful as measures of "pluralism" in a multi-dimensional trait space.

The generalization to mean  $S_n$  as presented here in 9 is different from that given by Laakso and Taagepera (1979). In both cases the core expression is  $\Sigma (S_i)^n$ , which is dimensionally different from  $S$ , e.g., if  $S$  is in dollars, then  $S_i^2$  is in dollars squared. In order to obtain an effective size, the expression must be reconverted to the original dimension. Laakso and Taagepera do it by taking the  $(n-1)^{\text{th}}$  root of  $\Sigma S_i^n / S$ ; i.e.,

$$S'_n = (\Sigma S_i^n / S)^{\frac{1}{n-1}} \quad [12]$$

As seen in Table 1, in this system  $S'_0$  is the arithmetical mean,  $S'_1$  is closely connected to entropy,<sup>8</sup> and  $S'_2$  is the grass-roots effective size.

Another way to correct the dimensionality is to divide  $\Sigma S_i^n$  by the similar expression involving  $n-1$  instead of  $n$ . This is the approach followed here (equation 9). As seen in Table 1, this approach also yields the arithmetic mean, but for  $n = 1$  rather than  $n = 0$  (which now yields the harmonic mean, while no entropy-connected expression emerges). The two normalization methods yield the same result for  $n = 2$  (the grass-roots effective size), and

TABLE 1  
Generalized Effective Size for Some Simple Values of  $n^a$

$n$	Eq. 23 ( $\bar{S}'_n$ )	Eq. 3 ( $\bar{S}_n$ )
$-\infty$	$S_{\min}$	$S_{\min}$
-2	$\sqrt[3]{S/\sum (1/S_i)^2}$	$\sum (1/S_i)^2 / \sum (1/S_i)^3$
-1	$\sqrt{S/\sum (1/S_i)}$	$\sum (1/S_i) / \sum (1/S_i)^2$
0	$S/N$ Arithmetic Mean	$N/\sum (1/S_i)$ Harmonic Mean
1	$S/e^H$ "Entropic" Mean	$S/N$ Arithmetic Mean
2	$\sum S_i^2/S$ "Grass Roots" Mean	$\sum S_i^2/S$ "Grass Roots" Mean
3	$\sqrt{\sum S_i^3/S}$	$\sum S_i^3 / \sum S_i^2$
$\infty$	$S_{\max}$	$S_{\max}$

a.  $S_{\min}$  and  $S_{\max}$  designate, respectively the smallest and the largest of the components  $S_i$ .  $H$  = entropy (see Note 8). In all cases a corresponding effective number of components can be calculated from effective sizes shown, using  $N_n = S/S_n$ .

also for  $n$  tending toward plus or minus infinity where the effective size becomes the size of the smallest or the largest component, respectively). For all other values, the two methods yield different results. Members of each family of indices have various advantages and disadvantages, but we shall not attempt here an axiomatic characterization of properties desirable in such indices.

For each family of indices we may state the general results that

$$S = N_n \bar{S}_n \quad [13]$$

and

$$S = N'_n \bar{S}'_n, \quad [14]$$

where mean  $S_n$  and  $S'_n$  are as defined in expressions 9 and 12, respectively, and  $N_n$  and  $N'_n$  emerge from substituting values of mean  $S'_n$  and mean  $S'_n$  into expressions 13 and 14, respectively.

### *DENSITY AND CROWDING*

We may make use of the formula in expression 11 to distinguish between density and crowding. Density is customarily defined simply as

$$D = \frac{S}{A}, \quad [15]$$

where  $A$  is some specified area, and  $S$  is total population in that area. If we consider  $N$  distinct territorial units, each with population  $S_i$ :  $\sum S_i = S$ , in an area  $A_i$ :  $\sum A_i = A$ , we have

$$D = \frac{S}{A} = \frac{\sum S_i}{\sum A_i} = \frac{S}{A}. \quad [16]$$

Let us define  $D = \text{mean } D_1$ . Thus,

$$\bar{D}_1 = \frac{N_1 \bar{S}_1}{A} = D \quad [17]$$

It is apparent that mean  $D_1$  is *independent of the distribution of population across territorial units*. While this may be a useful property for many purposes, it fails to pick up on the intuitive notion that crowding is related somehow to the number of people in proximity to one another.

We might think that we could find indices without the property of distribution independence by looking at the class of indices

$$\bar{D}_n = \frac{N_n \bar{S}_n}{A} . \quad [18]$$

However, it is apparent from expression 7 that mean  $D_n = D$  for all  $n$ . Thus, we must turn in a different direction.

A natural direction to look is toward the notion of neighborhood density. By an individual's neighborhood, we shall mean some sphere surrounding him. By fixing the radius of such a sphere, we can identify for each individual how many others are in his neighborhood. If we now let  $q_j$  be the number of individuals in the neighborhood of individual  $j$ , then for a fixed radius we may define

$$\bar{D}_r = \frac{\sum_{j=1}^S q_j}{\pi r^2 \sum_{i=1}^N S_i} = \frac{\sum_{j=1}^S q_j}{\pi r^2 S} . \quad [19]$$

As defined by expression 19, mean  $D_r$  is a natural candidate for an index of crowding. It is insensitive to the amount of space being occupied (e.g., it would be unchanged if an unpopulated desert area were to be annexed) but highly sensitive to the spatial distribution of individuals.

We may partition the population into  $N$  subsets,  $P_1, P_2, \dots, P_N$ , such that

$$\sum_{i=1}^N P_i = S,$$

and such that each of the members of the subset  $P_i$  perceive exactly  $q_i$  individuals in their neighborhood, and re-express expression 19 as

$$\bar{D}_r = \frac{\sum_{i=1}^N q_i P_i}{\pi r^2 \sum_{i=1}^N P_i}. \quad [20]$$

Note that, except for a normalizing factor, this expression can be shown to be identical in form to expression 11, and is analogous to expression 8.

If we were to assume that, because of rough symmetry considerations, wherever individual  $j$  sees  $q_{j-1}$  individuals in his neighborhood,  $q_j$  individuals also have neighborhoods with  $q_{j-1}$  members, we may specify a partition in which  $q_j = P_j$ .

With this simplifying approximation we may rework expression 20 as

$$\bar{D}_r \cong \frac{\sum_{i=1}^N q_i q_{i-1}}{\pi r^2 \sum_{i=1}^N q_i} \cong \frac{\sum_{i=1}^N q_i^2}{\pi r^2 \sum_{i=1}^N q_i}. \quad [21]$$

Note that, except for a normalizing factor, this expression is identical to that in expression 1.<sup>9</sup>

It is clear that mean  $D_r$  varies with  $r$ . At one extreme, we may let  $\pi r^2 = A$ , in which case

$$\bar{D} \cong \frac{S}{A} = D. \quad [22]$$

As  $r$  is reduced, mean  $D_r$  first increases, since empty regions are neglected; but eventually, the decrease in  $r$  starts reducing mean  $D_r$  so that at another extreme,  $r = 0$  leads to mean  $D_r = 0$ .

Since the above presentation has been a rather abstract one, let us turn to a simple example to demonstrate the differences between density and our crowding measure.

Let us consider two territories of equal size and equal population, and hence of equal density. In the first there are six individuals spaced as are the inner six points in Figure 1. These are each at a distance  $d$  from the center of the space and each point is at that same distance  $d$  from its two nearest neighbors. In the second there are six individuals spaced as are the outer six points in Figure 1. For the outer six points, each is at a distance  $2d$  from the center of the space and each point is at the same distance  $2d$  from its two nearest neighbors. Both territories have the same density, yet clearly they differ in their crowdedness.

Using the measure of crowding (mean  $D_r$ ) given in expression 20, it is easy to see that when  $r = d$ , the crowdedness of the inner set of points is  $2/\pi r^2$ , while that of the outer set of points is 0. Similarly, some simple geometry shows that when  $r = \sqrt{3}d$ , the crowdedness of the inner set of points is  $4/3\pi r^2$ , while that of the outer set of points remains zero. If  $r = 2d$ , the crowdedness of the inner set of points falls to  $5/4\pi r^2$ , while that of the outer set rises to  $2/4\pi r^2 = 1/2\pi r^2$ . If  $r = 2\sqrt{3}d$ , the crowdedness of the inner set of points falls to  $5/12\pi r^2$ , while that of the outer set of points falls to  $4/12\pi r^2$ . If  $r = 4d$ , the crowdedness of both sets is identical and equals  $5/16\pi r^2$ . Finally, if  $r > 2d$ , each set has the same crowdedness, which is given by  $5/\pi d^2$ .

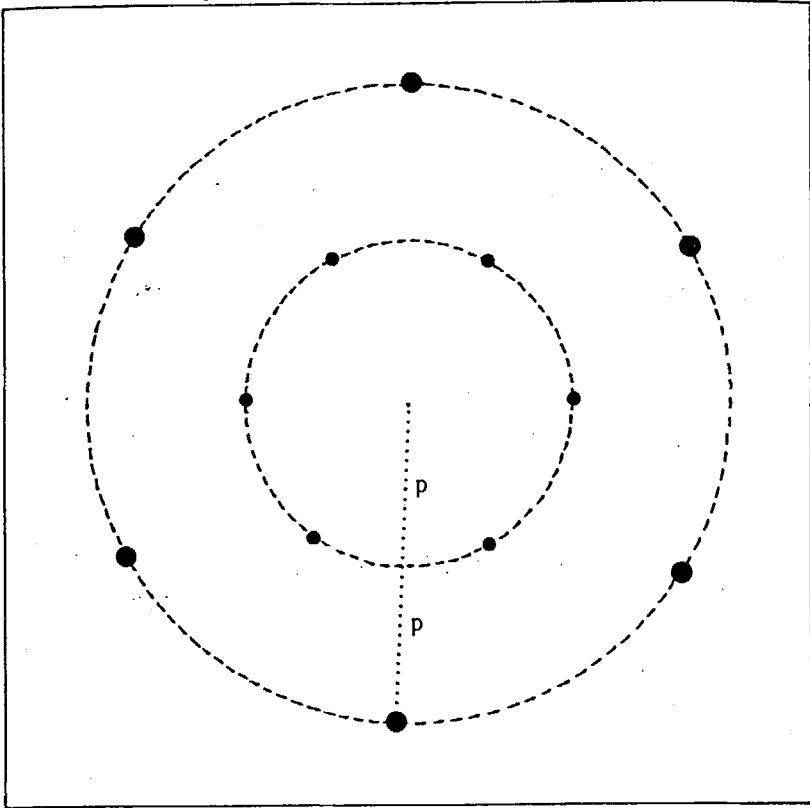


Figure 1: Two Dispersions of Population with Equal Density but Unequal Crowding

Several lessons emerge from this example. First, for  $r$  sufficiently large, two sets with equal density will be equally crowded as well. Second, for all values of  $r \leq 4d$ , the inner set is more crowded than the outer set. Third, for the special distributions given, the maximum values of crowdedness occur at  $d = r$  and  $d = 2r$ , respectively. Fourth, for the spatial arrays given, crowdedness does not monotonically decline with increasing  $r$ . For example, for the inner set of points, crowdedness is  $2/\pi r^2$  for  $d = r$ ,  $1/\pi r^2$  for  $d = \sqrt{2} r$ , but  $4/3\pi r^2$  for  $d = \sqrt{3} r$ .



We believe that maximum mean  $D_r$  is a strong candidate for a "natural"  $r$  value to use in a definition of crowdedness. By this definition the inner set of points in Figure 1 is four times as crowded as the outer set. Since the two sets of points are located on circles of radius  $d$  and  $2d$ , respectively, i.e., they have areas  $\pi d^2$  and  $4\pi d^2$ , respectively, this seems a plausible estimate of their relative crowdedness.<sup>10</sup>

### CONCLUSIONS

We have shown how the two seemingly distinct issues raised by Feld and Grofman (1977) and Laakso and Taagepera (1979) are conceptually intertwined. We have provided two families of indices of concentration certain of whose members may be particularly useful in a variety of applications, e.g., as measures of "effective size" and "effective number of components" in systems with components of unequal sizes and/or in which the exact size or number of all components is not known with certainty and as measures of "crowding." We hope our work will be useful in linking seemingly disparate ideas and measures into a common mathematical framework.

### NOTES

1. In particular, Feld and Grofman (1977) show that

$$\bar{S}_2 = \bar{S}_1 \left( 1 + \frac{\sigma^2}{\mu^2} \right).$$

Other implications of what Feld and Grofman refer to as the "class size paradox" are discussed in Feld and Grofman (1980).

2. This same point was made in an unpublished work by Grofman (1974).
3.  $HH$  has a maximum value of 1. As long as  $N$  is finite,  $N_2$  will also be finite.

4. Just as the same HH value may be obtained for different sets of  $P_i$  values, different sets of  $P_i$  values may yield the same  $N_2$  value. The value of treating different cases as equivalent is the same for both indices. However, under certain special circumstances, we may wish to consider distributional properties which these indices ignore. A useful case in point has been provided by an anonymous reviewer. The same value  $N_2 = 3.6$ , as obtained in our example, would also result (for the same total of 12.85 million votes) if there were one party with an absolute majority of 6.5 million, and ten other parties, each with 0.635 million votes. A single index is not sufficient to characterize the whole situation in such skewed cases, and an additional measure of "imbalance" (see Taagepera, 1980) is needed. The issue is somewhat analogous to that of sometimes having to complement standard deviation with skewedness and kurtosis. However, in the absence of extreme "imbalance" in the distribution, we may usefully treat systems which give rise to identical HH (or  $N_2$ ) values as if they were identical in their "concentration" (or "effective size").

5. We had a special reason for discussing the French communes rather than, say, the U.S. cities and towns: the realization that the Feld-Grofman (1977) and Laakso-Taagepera (1979) approaches are equivalent was triggered by a talk on French communal politics that Professor Jeanne Becquart-Leclercq gave at the University of California, Irvine (4/10/78).

6. It might be noted that the U.N. Security Council, designed to represent all world major powers and regions, has about  $N_2$  seats (11 until 1966 and 15 later on). We suspect that  $N_2$  will prove useful in picking sample sizes so as to yield "representative" sample where components are of unequal size. This issue we hope to pursue in further research.

7. Since mean  $S_0$  is extremely sensitive to the smallness of the smallest components, it might be a useful measure of inequality and deprivation. However, this point will not be elaborated in the present paper.

8. Rather than making use of the HH index to define an effective number of components, one could start with the system's entropy (H):

$$H = - \sum_{i=1}^N P_i \ln P_i.$$

Entropy is the central concept connecting social sciences to thermodynamics and information theory (see, e.g., Theil, 1967; Kittel, 1958). Then the exponential of entropy is an effective number ( $N_e$ ) of equal components which would yield the same entropy as does the actual system with unequal components:

$$N_e = e^H = \exp \left( - \sum_{i=1}^N P_i \ln P_i \right).$$

The values obtained tend to be somewhat higher than those of  $N_2$ . For the 1976 U.S. car industry,  $N_2 = 2.3$ , but  $N_e = 2.8$ . For the hypothetical party system discussed above,  $N_2 =$

3.6, but  $N_e = 4.4$ . Laakso and Taagepera (1979) gave a generalized expression which includes both  $N_e$  and  $N_2$  as special cases. They calculated both values for more than a hundred West European elections, and came to prefer  $N_2 = 1/HH$ , because the entropy-based  $N_e$  depended too much on small components, the sizes of which are often not known with any precision.

9. We might wish to define a family of indices for density analogous to that given in equation 9:

$$\bar{D}_r^{(n)} \cong \frac{\sum q_i^n}{\pi_r^2 \sum q_i^{n-1}},$$

but we shall not pursue that direction further here.

10. An alternative approach, however, might be to look at average crowding, defined perhaps as  $\int$  mean  $D_r$ . We suspect that, for most cases of interest the integral will be well behaved. We shall not, however, pursue this direction here.

## REFERENCES

- ALLISON, P. D. (1978) "Measures of inequality." *Amer. Soc. Rev.* 43 (December): 865-880.
- BLAU, P. M. (1977) *Inequality and Heterogeneity*. New York: Macmillan.
- CBS (1978) *The CBS Almanac* 1978.
- FELD, S. and B. GROFMAN (1980) "The class size paradox and conflict of interest among faculty, students, and administration." *Frontiers of Economics* 3.
- (1977) "Variation in class size, the class size paradox and some consequences for students." *Research in Higher Education* 6: 215-222.
- GIBBS, J. P. and W. T. MARTIN (1962) "Urbanization, technology, and division of labor: international patterns." *Amer. Soc. Rev.* 27 (October): 667-677.
- GREENBERG, J. (1956) "The measurement of linguistic diversity." *Language* 32: 109-115.
- GROFMAN, B. (1974) "Indices of representation and distortion." SUNY, Stony Brook. (unpublished)
- HERFINDAHL, O. C. (1950) "Concentration in the steel industry." Ph.D. dissertation, Columbia University.
- HIRSCHMAN, A. O. (1945) *National Power and Structure of Foreign Trade*. Berkeley: Univ. of California Press.
- KITTEL, C. (1958) *Elementary Statistical Physics*. New York: John Wiley.
- LAAKSO, M. and R. TAAGEPERA (1979) "'Effective' number of parties: a measure with application to West Europe." *Comparative Pol. Studies* 12 (April): 3-27.

- (1978) "Proportional representation in Scandinavia: implications for Finland." *Scandinavian Pol. Studies* 1: 43-60.
- LIEBERSON, S. (1969) "Measuring population diversity." *Amer. Soc. Rev.* 34 (December): 850-862.
- RAE, D. W. and M. TAYLOR (1970) *The Analysis of Political Cleavages*. New Haven, CT: Yale Univ. Press.
- TAAGEPERA, R. (1979) "Inequality, concentration, imbalance." *Pol. Methodology* 6: 275-291.
- and J. L. RAY (1977) "A generalized index of concentration." *Soc. Methods & Research* 5 (February): 367-384.
- TAYLOR, C. L. and M. C. HUDSON (1972) *World Handbook of Political and Social Indicators*. New Haven, CT: Yale University Press.
- THEIL, H. (1967) *Economics and Information Theory*. Amsterdam: North Holland.

*Bernard Grofman is Professor of Political Science and Social Psychology, School of Social Sciences, University of California, Irvine. He is a specialist in mathematical models of collective decision-making, with over two dozen published articles on topics such as jury verdict choice, voter turnout, coalition formation models, and the political consequences of electoral laws.*

*Rein Taagepera is presently Professor of Social Science at the University of California, Irvine. His recent papers (those published since 1979, or in press) deal with arms race and world population growth models; growth curves of historical empires and of Western civilization; seat-vote proportionality under various electoral rules; measurement of inequality, concentration and imbalance; and Sovietization, collectivization and guerrilla resistance in postwar Estonia.*