Title
A Generalized Index of Concentration

Journal
Sociological Methods & Research, 5(3)

ISSN
0049-1241

Authors
Taagepera, Rein
Ray, James Lee

Publication Date
1977-02-01

DOI
10.1177/004912417700500306

Peer reviewed

A generalized nth power index of concentration is defined; for n = 1 it yields the entropy-based "relative redundancy," and for n = 2 it yields the Ray and Singer (1973) concentration coefficient. Relations are established with a dozen other indices of concentration, diversity, fragmentation, differentiation, and inequality used in social science literature. Explicit conversion formulas are given for these often apparently unrelated indices. For multicomponent systems, equivalent two-component systems are defined.

A GENERALIZED INDEX OF CONCENTRATION

REIN TAAGEPERA
University of California, Irvine
JAMES LEE RAY
University of New Mexico

Devising indices to reflect the related concepts of concentration and of inequality is a task which has occupied the time and talents of a sizable number of scholars. This is, perhaps, not surprising, since inequality is a concept which plays an important role in several and diverse areas of social science, be it economics (where the focus is, usually, on the inequality of the distribution of wealth), political science (where interest centers on the distribution of power in societies, legislative bodies, or the international system), or sociology (where the diversity of races, ethnic groups, or classes in a society receives the most attention).

The result of this wide interest in the concept of inequality and persistent index-construction efforts is, naturally enough, a large number of indices. The definitions of a number of them are given in Table 1. Indeed, anyone who decides to use a measure of concentration or inequality in his work may find an embarrassment of riches if he tries to select among these alternatives.

A review of the literature suggests that most social scientists are likely to select the Gini index. This index, as Alker (1965:

SOCIOLOGICAL METHODS & RESEARCH, Vol. 5 No. 3, February 1977

41) explains, "sums for each individual in the population the difference between where he is on the Lorenz curve, and where he would be expected to be in the case of democratic equality." The Gini index is basically reasonable, but, as one of us has shown in a previous article (Ray and Singer, 1973), it has some characteristics which make it less than ideal for some purposes. Most importantly, if one uses Alker's (1965) formula for approximations in case of discrete distributions (which is necessary because the Gini index was originally devised for continuous distributions), the index has an upper limit of $(1 - 1/N)$ where N is the number of components. If one is comparing distributions among systems with small numbers of components, this upper limit can exert a distorting effect on the index scores. A distribution between a two-component entity may receive a much lower inequality score than a distribution

TABLE 1
Conversion Formulas for Some Concentration and Inequality Indices

DEFINITION OF TERMS

T = total number of items to be shared.

N = number of categories or components which share the items.

T_i = number of items belonging to the i-th category.

P_i = fractional share of the i-th category; $P_i = T_i/T$; $\Sigma P_i = 1$.

P_1 = fractional share of the largest category.

For indices involving rank, i indicates rank; in other cases the order in which the categories are taken does not matter.

Unless otherwise stated, Σ indicates summation over all categories from $i = 1$ to $i = N$.

	Definition of Index	Conversion Formuli
GENERALIZED INDICES OF DEGREE n :		
Non-normalized concentration index	$B_n = \frac{\Sigma P_i^{\,n} - N^{1-n}}{1 - N^{1-n}}$	$B_n = (C_n)^n$
Normalized concentration index	$C_n = (B_n)^{1/n}$	
Normalized average	$A_n = \frac{(\Sigma P_i^{\,n}/N)^{1/n} - \frac{1}{N}}{(\frac{1}{N})^{1/n} - \frac{1}{N}}$	$A_n = \frac{\left[(C_n)^n (1 - N^{1-n}) + N^{1-n}\right]^{1/n} - N^{1/n-1}}{1 - N^{1/n-1}}$
INDICES CONVERTIBLE INTO C_1 :		
Entropy	$H = -\Sigma P_i \ln P_i$	$H = \ln N(1 - C_1)$
Redundancy	$R = \ln N - H$	$R = C_1 \ln N$
Relative redundancy	$RR = \frac{\ln N - H}{\ln N}$	$RR = C_1$

TABLE 1 (Continued)

INDICES CONVERTIBLE INTO C_2 :		
Herfindahl-Hirschman index of industrial concentration	$HH = \Sigma P_i^2$	$HH = C_2^2(1 - \frac{1}{N}) + \frac{1}{N}$
Greenberg linguistic diversity index	$G = HH$	$G = C_2^2(1 - \frac{1}{N}) + \frac{1}{N}$
Michaely geographic concentration %	$M = 100\sqrt{HH}$	$M = 100\sqrt{C_2^2(1 - \frac{1}{N}) + \frac{1}{N}}$
Ray-Singer concentration coefficient	$CON = \sqrt{\frac{HH - \frac{1}{N}}{1 - \frac{1}{N}}}$	$CON = C_2$
Major component of equivalent two-component system		$P_1 = (C_2 + 1)/2$
Amemiya index of economic differentiation	$IED = \Sigma\frac{N}{N - 1}(P_i - \frac{1}{N})^2$	$IED = C_2^2$
Mueller index of qualitative variation	$IQV = \frac{\Sigma\Sigma T_i T_j}{\frac{N(N-1)}{2}(\frac{T}{N})^2}$	$IQV = 1 - C_2^2$
Janda index of party articulation	J	$J = C_2$
Rae and Taylor index of fragmentation	$RT = 1 - \frac{\Sigma T_i(T_i - 1)}{T(T-1)}$	$RT = \frac{(1 - \frac{1}{N})(1 - C_2^2)}{1 - \frac{1}{T}}$
Labovitz & Gibbs degree of division of labor	$D = 1 - \frac{(\Sigma T_i^2)/(\Sigma T_i)^2}{1 - \frac{1}{N}}$	$D = 1 - C_2^2$
OTHER INDICES NOT INVOLVING RANK		
Horvath CCI	$CCI = P_1 + \sum_{i=2}^{N} P_i^2 [1 + (1 - P_i)]$	$CCI = f(C_2, C_3, P_1)$
Schutz coefficient	$S = \frac{1}{2}\Sigma\sqrt{(P_i - \frac{1}{N})^2}$	None
INDICES INVOLVING RANK		
Gini index	$G = 1 + \frac{1}{N} - (\frac{2}{N})\Sigma iP_i$	$G = 1 - \frac{1}{N\,TH}$
Hall & Tideman index	$TH = \frac{1}{2\Sigma iP_i - 1}$	$TH = \frac{1}{N(1 - G)}$

among a five-component entity simply because the upper limit of the index is .50 in the first case, but .80 in the second case. This problem may seem obvious, or even trivial, but it has been ignored, with unfortunate results (e.g., by Starr, 1972), and it seemed worth pointing out.

The scholar in search of an index might turn to the Schutz coefficient (Alker, 1965). This measure is based on the slope of the Lorenz curve, and in effect reflects how close the slope of the line below the equal share point is to zero, or how close the slope of the line above the equal share point is to infinity. This would be a poor choice, however, since this index also has an upper limit of $(1 - 1/N)$, as shown by Ray and Singer (1973).

If our scholar is inclined to continue his search, he is likely to find that economists have devised a series of indices of industrial concentration, i.e., inequality of the distribution of shares of the market among the firms in a given industry. One of the best known is the Herfindahl-Hirschman (HH) index of industrial concentration (Herfindahl, 1950), which equals the sum of the squares of each component's percentage share. This measure has been refined at least a couple of times to meet various specialized needs, e.g., by Hall and Tideman (1967) and by Horvath (1970)—see definitions in Table 1. All of these are worth considering if one is interested in industrial concentration. But, again, they may not be appropriate in every case when one is interested in some other kind of inequality. HH, for example, has a *lower* limit of $(1 - 1/N)$, and the scores of this index are greatly affected by a change in N—especially, of course, when N is small.

This problem, and others, led Ray and Singer (1973) to create CON, defined as

$$\mathrm{CON} = \sqrt{\frac{\sum_{i=1}^{N} P_i^2 - \frac{1}{N}}{1 - \frac{1}{N}}}\ , \qquad [1]$$

where P_i equals the fractional share of the ith category. CON is essentially a normalized coefficient of variation (cf. Theil, 1967: 125). One of the advantages of such a measure is that its range is always 0 to 1, no matter what N is. The process of creating that measure lead to the discovery that there is a rather lengthy list of concepts, all of which are related to but different from inequality, and all of which have been measured with indices which, like CON, are based on ΣP_i^2. Greenberg (1956), for example, presents such an index to measure linguistic diversity. Rae and Taylor (1970) have used a similar index to measure fragmentation, or party concentration in parliaments and other legislative bodies. Michaely (1962) uses such an index to measure geographic concentration of transactions. Other important examples include Lieberson's (1969) index of population diversity, Bachi's (1956) and Simpson's (1949) indices of diversity, the Bell (1954) index of ecological segregation, and the Gibbs and Martin (1962) index of diversification in industry.

There is an additional set of concepts, the indices for which are not only based on ΣP_i^2, but which are also virtually identical to CON. For example, Amemiya's (1963) index of economic differentiation is equal to CON^2. Labovitz and Gibbs (1964) present a measure of the division of labor (D) which is equal to $(1 - CON^2)$. Exactly equivalent to D is the Mueller et al. (1970) index of qualitative variation (IQV). Finally, Janda's (1971) index of party articulation is exactly equivalent to CON.

By now it should be obvious that there is a large number of concepts which share similarities with the concept of equality (and, of course, with the opposite concepts of inequality and concentration). Even those concepts discussed above (and in Ray and Singer, 1973) do not, however, constitute the full list. Another concept related to concentration and quality is entropy (H), which we will discuss in more detail.

AN ENTROPY-BASED INDEX OF CONCENTRATION

The concept of entropy (H) occupies a central place in thermodynamics and statistical physics (see, e.g., Kittel, 1958) as well as in information and communication theory (see, e.g., Yaglom and Yaglom, 1969). Its use in measurement of economic inequality has been discussed in detail by Theil (1967), who has also used it (1969) to solve political problems involving the distribution of seats and votes among various parties. Since few concepts manage to be useful both in physical and social sciences, entropy is a concept of great and possibly unsurpassed generality in sciences.

If there are N possible states in which a system can be, and if P_i is the probability that the system will be found to be in state i, then entropy is defined as

$$H = -\sum_{i=1}^{N} P_i \ln P_i. \qquad [2]$$

For our purposes P_i represents the fractional share of the ith component or category. If concentration of property is complete, one of the P_i is unity and all the others are zero; the corresponding value of entropy can be shown to be zero. On the other extreme, if there is complete equality, $P_i = 1/N$ for all the N components; entropy then reaches its maximum value, which is $\ln N$.

If we want to have an index ranging from 0 to 1, we will have to normalize H by dividing it by its maximum value. However, this "relative entropy" $H/\ln N$ is zero for complete concentration and unity for complete equality; it is, hence, an index of uniformity or equality. In order to have an index of concentration or inequality we merely have to subtract the relative entropy from one. The resulting concentration index is often called "relative redundancy" (RR) in communication theory. In terms of entropy,

$$RR = \frac{\ln N - H}{\ln N}. \qquad [3]$$

In terms of the fractional shares of the N components,

$$RR = \frac{\ln N + \sum_{i=1}^{N} P_i \ln P_i}{\ln N}. \qquad [4]$$

Theil (1967: 92n.) prefers to use the nonnormalized redundancy ($\ln N - H$) because it can more easily be decomposed into individual-to-subgroup and subgroup-to-group redundancies. The upper limit of redundancy ($\ln N$) increases with the number of components. It can well be argued that concentration of power into the hands of one person is greater in the case of a total population of one million than in the case of a total population of 10 people; redundancies would be 13.8 and 2.3, respectively. But for purposes of comparison there are also advantages in using relative redundancy where the maximum possible concentration within a system leads to index value "one" irrespective of the system size.

Compared to the Ray and Singer concentration index CON [1] which also varies from 0 to 1, RR does not depend as heavily on the share of the largest component. (This is so because the product of P_i by itself has been replaced by the product of P_i and its natural logarithm.)

While many other inequality and concentration coefficients can be easily converted to CON because they have the same core expression ΣP_i^2, the relative redundancy cannot, because its very summational core is different: $\Sigma P_i \ln P_i$. The whole appearance of the two indices appears fundamentally different, with CON containing a square root sign, and with RR involving logarithms rather than powers of N and of P_i. However, it will be seen that both CON and RR are special cases of a more general concentration index, which will be presented next.

A GENERALIZED INDEX OF CONCENTRATION

Apart from the conceptually satisfying interpretation of CON as normalized standard deviation, the power index 2 in ΣP_i^2 is not critical for defining a concentration index—we would also obtain a workable index if we used a power index of 3 or of 1.5 instead of 2. Normalization terms would of course be different.

Consider the generalized summational core ΣP_i^n where n is any positive number. For complete concentration ($P_1 = 1$, and $P_i = 0$ for $i \neq 1$) its value is 1. For complete equality of all components ($P_i = 1/N$ for all i) it reaches a minimum value of $[N(1/N)^n] = N^{1-n}$. In order to have a zero value for complete equality we have to subtract N^{1-n} from the core. In order to maintain the maximum value of 1 for complete concentration we must divide by $(1 - N^{1-n})$. The result is

$$B_n = \frac{\sum_{i=1}^{N} P_i^n - N^{1-n}}{1 - N^{1-n}} . \qquad [5]$$

Since we first raised all the P_i's to the nth power, we might want to carry out the inverse operation (i.e., taking the nth root) on the normalized expression, out of a general sense of symmetry. The generalized nth degree concentration index is then

$$C_n = \left[\frac{\sum_{i=1}^{N} P_i^n - N^{1-n}}{1 - N^{1-n}} \right]^{\frac{1}{n}} . \qquad [6]$$

Obviously, for $n = 2$, we reobtain the expression in equation 1, so that $CON = C_2$. A less obvious finding is that the relative redundancy is generated by the same general expression when n tends toward the value 1: $RR = C_1$.[1]

Another possible way to carry out the inverse operation would be to do so on the core, before normalizing:

$$A_n = \frac{\left(\sum_{i=1}^{N} P_i^n \right)^{1/n} - N^{(1-n)/n}}{1 - N^{(1-n)/n}} = \frac{\left(\sum_{i=1}^{N} P_i^n/N \right)^{1/n} - \frac{1}{N}}{(1/N)^{1/n} - \frac{1}{N}} . \qquad [7]$$

The expression $(\Sigma\ P_i^n/N)^{1/n}$ is a generalized nth power average of the N quantities P_i. With $n = 1$, it yields the usual arithmetic average. With $n = 2$ it yields the "root mean square" of P_i:

$$RMS = \sqrt{\sum_{i=1}^{N} P_i^2/N} \qquad [8]$$

which is widely used in physics, e.g., for diffusion of gases and for alternating currents (Semat, 1957: 312, 592). Thus, A_n is a normalized generalized average.

In order to compare the characteristics and merits of C_1, of C_2, and of other possible C_n, complete $C_n(n)$ curves were calculated and plotted for a variety of distributions of shares.

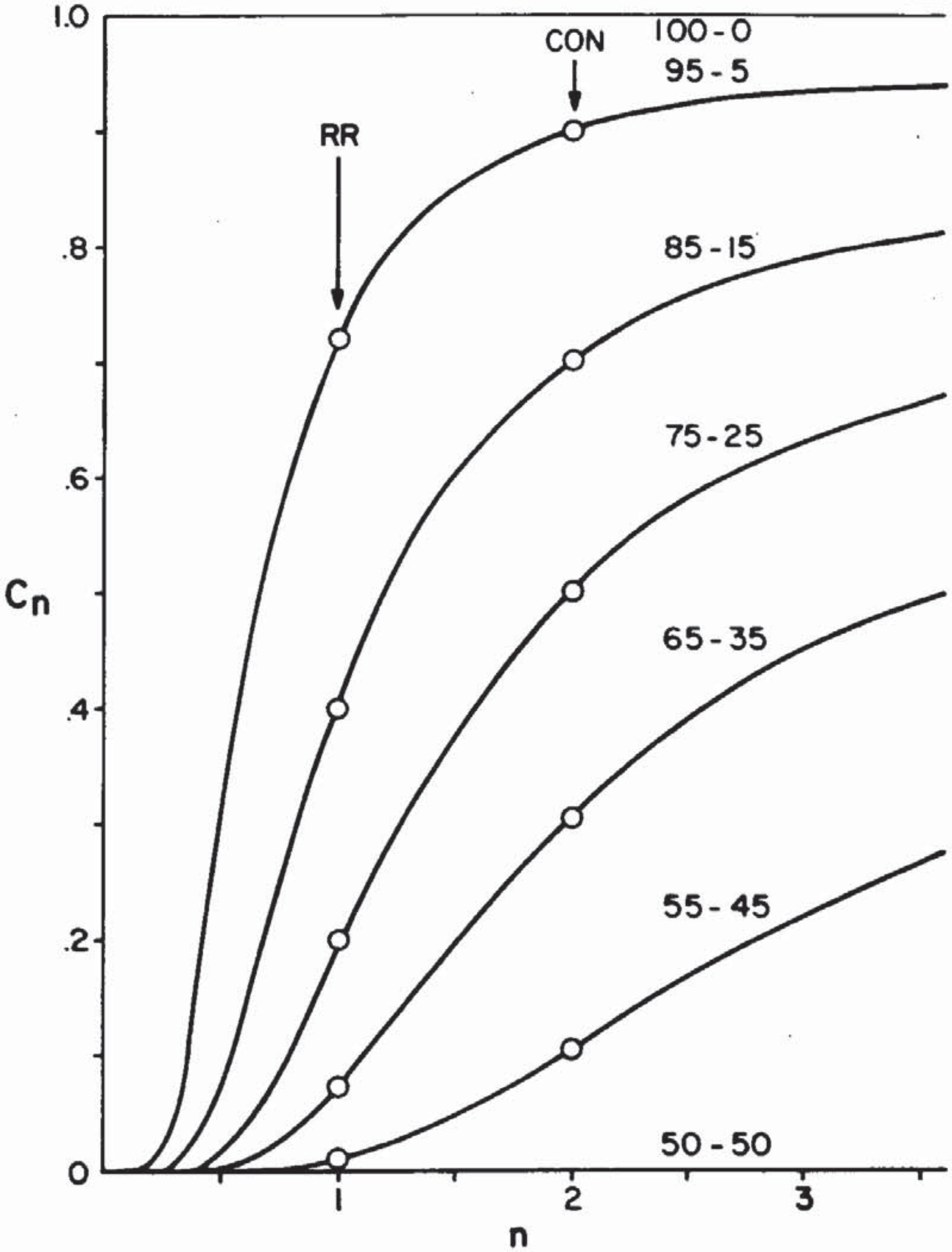

Figure 1: Variation of the generalized concentration index C_n with n, for various concentration levels (in percent shares) of a two-component system. Note locations of the entropy-based Relative Redundancy and of the variation coefficient-based CON.

The same was done for the intermediary expression B_n and for the alternative expression A_n.

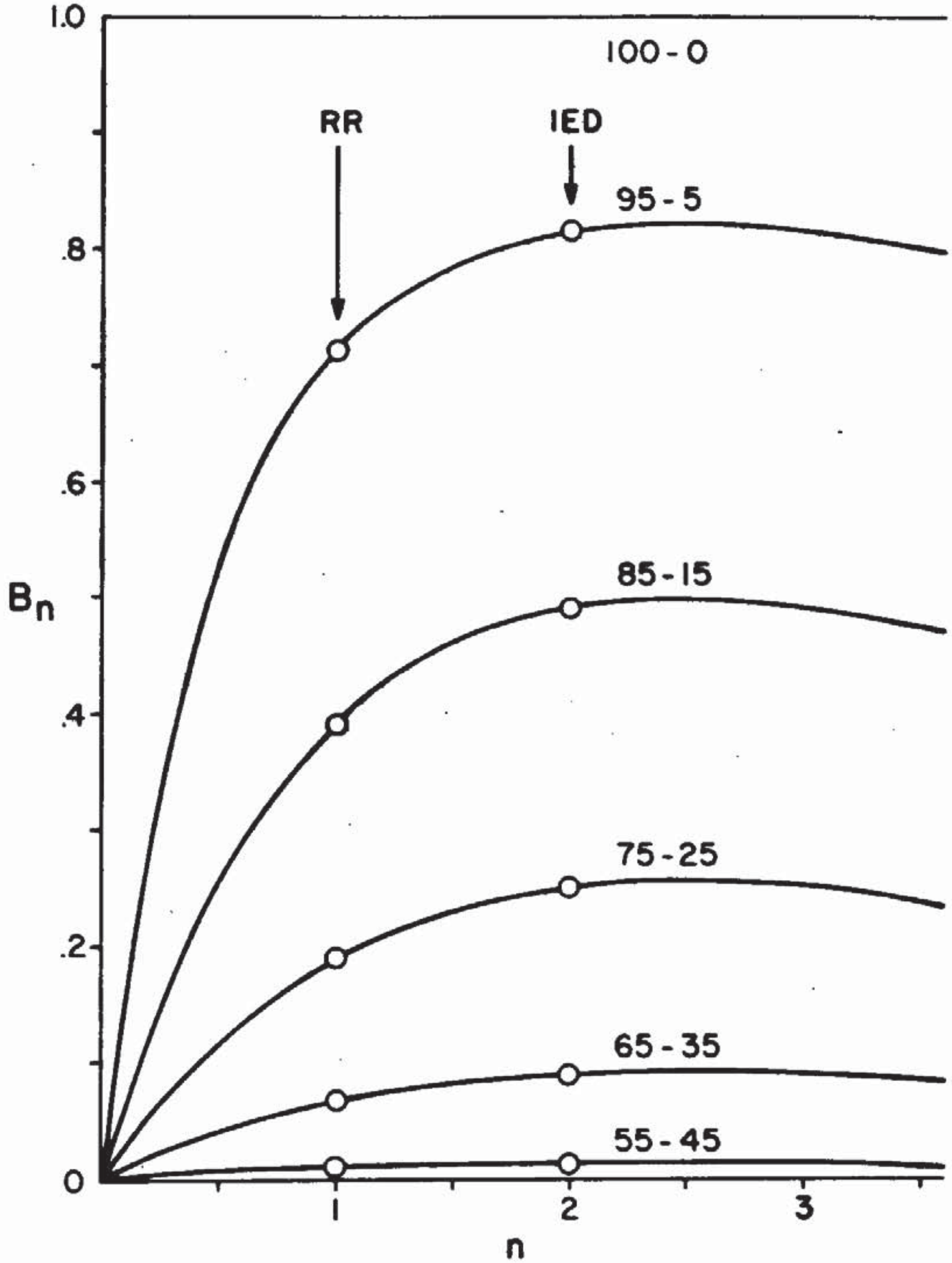

Figure 2: Variation of the intermediary index B_n with n, for various concentration levels of a two-component system. Note locations of Relative Redundancy and of Amemiya Index of Economic Differentiation.

Consider first the simplest possible structure of wealth distribution—the one involving two components only: N = 2

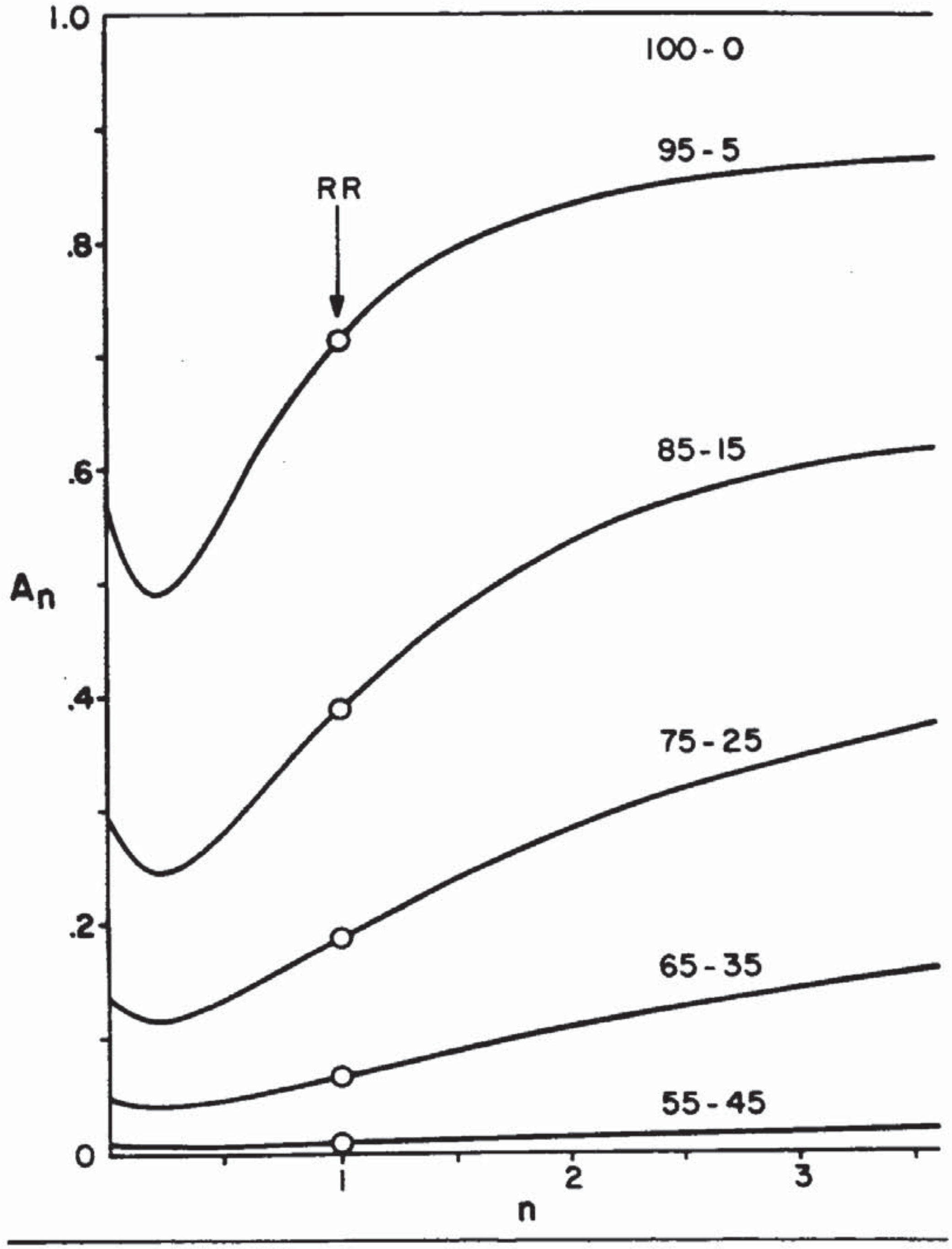

Figure 3: Variation of the normalized generalized average A_n with n, for various concentration levels of a two-component system.

and $P_1 + P_2 = 1$. The corresponding $C_n(n)$ curves for various power distributions are shown in Figure 1. The $B_n(n)$ curves are shown in Figure 2, while the $A_n(n)$ curves are shown in Figure 3. These patterns are quite typical, for any number N of components.

For n = 0 we always have $C_0 = 0$. As n tends to infinity, C_n tends to the value P_1 of the largest of the shares P_i. In most cases the slope of the $C_n(n)$ curve stays positive (as in Figure 1), so that P_1 is the maximum value C_n takes. The only exception arises when the two largest components have equal or nearly equal shares. An example is shown in Figure 4, where the percent distribution 45-35-20-0-0 yields a peak value C_n = 0.4744 at n = 3.3 while the value at infinite n is $P_1 = 0.45$. For a ten-component system where five shares are equal to 20% and the other five are 0%, the peak is around $C_{1.8} = 0.335$. In all cases C_2 is larger than C_1.

The curve for intermediary index B_n usually starts with B_0 = 0, reaches a peak at n less than 3, and gradually decreases again to 0 as n tends to infinity, as seen in Figure 2. An exception occurs when some of the components have perfectly zero shares; then $B_0 = N'/N - 1)$ where N' is the number of perfectly zero components. In the example shown in Figure 4 (percent distribution 45-35-20-0-0), the initial value $B_0 = 0.5$ is also the maximum value of B_n, but this is not always the case. Thus, for the distribution 70-20-10-0-0 the peak value is around $B_{0.5}$ = 0.515, while $B_0 = 0.5$. If the smallest components are ever so slightly larger than zero, then B_0 immediately drops to 0 (as seen in Figure 4, for distribution 45-34.99-19.99-0.01-0.01), while the rest of the $B_n(n)$ curve is not visibly altered. For N = 2 the peak of B_n occurs around n = 2.5, and it can easily be shown that $B_2 = B_3$. For a larger number of components the peak value always has been found to be at n less than 2, although we cannot prove that it must be so. The qualitative decision on whether to include or exclude zero or near-zero size components (45-35-20-0-0 versus 45-35-20, in Figure 4) strongly affects the outcome.

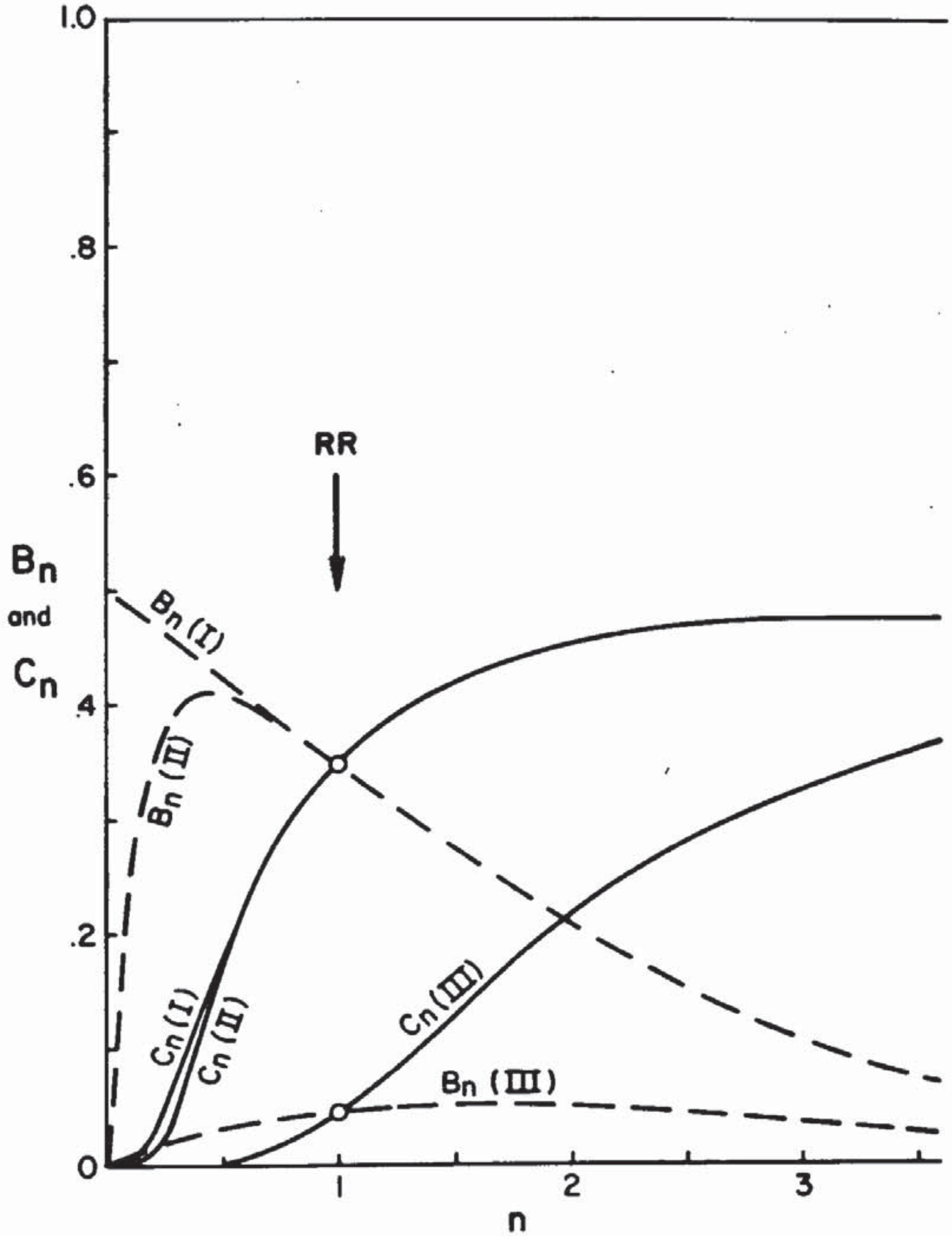

Figure 4: The $C_n(n)$ and $B_n(n)$ curves for three quantitatively similar but qualitatively different percent share configurations: (I) 45-35-20-0-0—several zero components; (II) 45-34.99-19.99-0.01-0.01—several very small components; (III) 45-35-20—zero or near-zero size components omitted. The C_n and B_n curves always cross at n = 1 where $C_1 = B_1$ = Relative Redundancy.

The curve for the normalized generalized average A_n starts from $A_0 = 1 - NG$ where G is the geometric average of all the P_i. As n increases, A_n usually decreases and then increases again, but patterns sometimes are more complex. As n tends to infinity, A_n always tends to $(NP_1 - 1)/(N - 1)$ where P_1 is the share of the largest component.

In order to describe inequality systems one could choose arbitrarily a certain value of n and one of the expressions C_n, B_n, or A_n. But it would be preferable to have a value of n dictated by some features of the very pattern in Figures 1 to 3, such as minima or maxima of curves. Unfortunately, the maximum of the $C_n(n)$ curve almost always depends on the largest share P_1 alone, while the maximum of the $B_n(n)$ curve depends almost uniquely on the share of the smallest components (cf. Figure 4). The minimum of the $A_n(n)$ curve also seems to be a poor indicator of the degree of inequality. Choosing arbitrarily the value n = 1 to describe the system would have the advantage of saving us a choice between the generalized expressions A_n, B_n, and C_n, since $A_1 = B_1 = C_1 =$ RR. Another possibility is to reduce every inequality system to a simpler system as described in the next section.

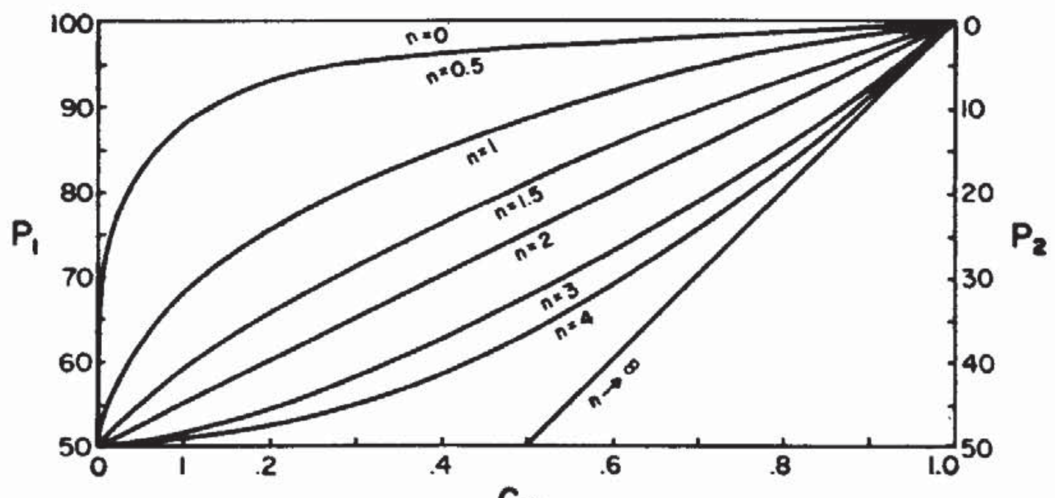

Figure 5: Equivalent two-component systems at various n values, for given C_n. P_1 is the major component and P_2 is the minor component of the two-component system (in percentage).

EQUIVALENT TWO-COMPONENT SYSTEMS (ETS)

If a system with three or more components has the same C_2 value as a two-component system, then we could say that these systems are equivalent on the level n = 2. Establishing such equivalent two-component systems (ETS) is made easy by the use of Figure 5, which shows the share P_1 of the major component of the two-component system which has a given value of C_n. (Figure 5 is just another way of looking at the data in Figure 1.) For example, the system 32-32-32-2-2 has C_2 = .365 and is thus seen to be equivalent to a 68-32 system on level n = 2, according to Figure 5. On the level n = 1 (C_1 = .223) it is equivalent to a 77-23 system. When the system contains many small components, the equivalent two-component systems on level n = 2 and n = 1 can differ considerably; thus, for the system 45-35-20-0-0 we would have C_2 = .435 leading to an ETS of 72-28, but C_1 = .348 leading to an ETS of 83-17. However, in most cases the ETS are more or less the same over a wide range of n while the values of C_n differ considerably, depending on n. Thus, the ETS might represent a fairly stable basis for comparing system concentrations. It would further have the appeal of direct visualization of its meaning.

A simple relation between C_n and the major component P_1 of the corresponding ETS exists for n = 2 and only for n = 2:

$$2P_1 = C_2 + 1 = \left(\frac{\Sigma P_i^2 - \frac{1}{N}}{1 - 1/N}\right)^{\frac{1}{2}} + 1. \qquad [9]$$

The minor component is of course simply $1 - P_1$. Because of this simple relation (which stands out in Figure 5 as a straight line) this definition for ETS might have some merit.

CONVERSION FORMULAS

Concentration and inequality are related concepts (see Theil, 1967: 128) but they are not quite identical. Consider two isolated continents, each of which has complete concentration of power in the hands of one country, so that $C_n = 1$ and the Gini index is close to one. When the two continents establish contacts and thus form a single system, concentration (as measured by C_n) drops markedly, while the Gini index of inequality does not change.

As shown in the introduction, a large variety of concentration and inequality indices (and also of equality and fragmentation indices) have been proposed and used. Table 1 shows a number of them, with definitions and, where possible, conversion formulas into the generalized C_n notation. Relationships with C_2 have been discussed in detail in Ray and Singer (1973) for the indices RT, M, IED and D. The same paper also developed the expression for the Gini index as used in Table 1.

In general, concentration and inequality coefficients fall into two broad categories.

(1) Indices where the summational core contains both the share sizes P_i and the rank index i. Two such indices are used–Gini and TH. They are easily convertible into each other (as shown in Table 1) but they cannot be converted into the C_n system. Before these indices can be calculated, the components must be rearranged in order of decreasing share size to determine the respective i values. These indices may reflect inequality more than concentration.

(2) Indices where the summational core contains P_i but not i. In these cases i is merely a label (rather than a rank indicator) and the components can be labeled and added in any order. Such indices can usually be converted into the C_n; the only exception is the Schutz coefficient. These indices may reflect concentration more than inequality.

Irrespective of the index used, we are facing the qualitative dilemma illustrated in Figure 4: whether to exclude or include very small-size categories. Was Austria in 1939-1944 a nonexisting state or a zero-power state? When calculating concentration of automotive industry, do you include workshops producing five specialty cars per year? These issues are far from being clear-cut, and results differ considerably depending on the total number N of components we choose to include.

CONCLUSIONS

We have formulated a generalized nth power index of concentration (C_n) and established its relations with most previously used indices. It cannot be said that any of the levels n is clearly preferable, but both C_1 and C_2 have certain desirable special properties. In general, higher levels of n yield higher concentration index values. The equivalent two-component system at n = 2 might offer a relatively stable and intuitively interpretable measure of inequality or concentration. Two qualitative problems remain to be solved: the treatment of small and zero-size components, and the relation between inequality and concentration.

NOTE

1. A proof of this result is available upon request from the first author.

REFERENCES

ALKER, H. (1965) Mathematics and Politics. New York: Macmillan.

AMEMIYA, E. C. (1963) "Measurement of economic differentiation." J. of Regional Sci. 5 (Summer): 84-87.

BACHI, R. (1956) "A statistical analysis of the revival of Hebrew in Israel," pp. 179-247 in R. Bachi (ed.) Scripta Hierosolymitana. Jerusalem: Magnus.

BELL, W. (1954) "A probability model for the measurement of ecological segregation." Social Forces 32 (May): 357-364.

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.

HALL, M. and N. TIDEMAN (1967) "Measures of concentration." J. of Amer. Statistical Assn. 62 (March): 162-168.

HERFINDAHL, O. C. (1950) "Concentration in the steel industry." Ph.D. dissertation, Columbia University.

HORVATH, J. (1970) "Suggestion for a comprehensive measure of concentration." Southern Econ. J. 36 (April): 446-452.

JANDA, K. (1971) "Conceptual equivalence and multiple indicators in the cross-national analysis of political parties." Prepared for the ISS/UNESCO/ECPR Workshop on Indicators of National Development, August.

KITTEL, C. (1958) Elementray Statistical Physics. New York: John Wiley.

LABOVITZ, S. and J. P. GIBBS (1964) "Urbanization, technology, and the division of labor: further evidence." Pacific Soc. Rev. 7 (Spring): 3-9.

LIEBERSON, S. (1969) "Measuring population diversity." Amer. Soc. Rev. 34 (December): 850-862.

MICHAELY, M. (1962) Concentration in International Trade. Amsterdam: North Holland.

MUELLER, J. H., K. F. SCHUESSLER, and H. L. COSTNER (1970) Statistical Reasoning in Sociology. Boston: Houghton Mifflin.

RAE, D. W. and M. TAYLOR (1970) The Analysis of Political Cleavages. New Haven, Conn.: Yale Univ. Press.

RAY, J. L. and J. D. SINGER (1973) "Measuring the concentration of power in the international system." Soc. Methods and Research 1 (May): 403-437.

SEMAT, H. (1957) Fundamentals of Physics. New York: Rinehart.

SIMPSON, E. H. (1949) "Measurement and diversity." Nature 163 (April): 688. 8.

STARR, H. (1972) War Coalitions. Lexington, Mass.: D. C. Heath.

THEIL, H. (1969) "The desired political entropy." Amer. Pol. Sci. Rev. 63 (June): 521-525.

--- (1967) Economics and Information Theory. Amsterdam: North Holland.

YAGLOM, A. M. and I. M. YAGLOM (1969) Probabilité et Information. Paris: Dunod.

Rein Taagepera is Associate Professor of Political Science at the University of California, Irvine. His published quantitative work deals with inequality of farm size distribution, growth curves of empires, world population growth models, arms races, representative assembly sizes, and effect of country size on trade/GNP ratio.

James Lee Ray is Assistant Professor of Political Science at the University of New Mexico. His interest in measurement problems stems largely from substantive concerns in the area of peace research. He is currently involved in research regarding "dependency" theory as it applies to inter-American relations.