Non-Profile Specific Measures of Social Homogeneity*

William V. Gehrlein

University of Delaware

Correspondence on this paper should be addressed to:

Dr. William V. Gehrlein
Department of Business Administration
University of Delaware
Newark, Delaware 19711
United States of America

* This research was supported by a grant from the National Science Foundation (United States)

Abstract

Social homogeneity refers to how much the preferences of individuals in a society tend to be alike. A number of studies have been conducted to determine whether or not a relationship exists between various measures of social homogeneity and the probability that a Condorcet winner exists. In this study, it is shown that a strong general relationship of this type does not exist for measures of social homogeneity which account only for the proportions of individuals with various preference structures. That is, for measures which account for these proportions but not for the preference structures themselves.

Introduction

Social homogeneity refers to how much the preferences of individuals in a society tend to be alike. We shall develop this notion in the context of an election on a set of three candidates. Each member of society is assumed to have some linear preference ranking on the candidates, so that no individuals have any feeling of indifference between pairs of candidates. For three candidates, {A, B and C}, there are six linear preference rankings given by:

$A > B > C$: p_1

$A > C > B$: p_2

$B > A > C$: p_3

$B > C > A$: p_4

$C > A > B$: p_5

$C > B > A$: p_6.

Here, $x > y$ denotes x is preferred to y and p_i is the probability that an individual selected at random from the society will have the associated preference ranking with $\sum_{i=1}^{6} p_i = 1$. In a specific voting situation, the preferences of n voters from the society are defined as a profile in terms of six n_i's. Here, n_i is the number of voters with the ith preference ranking so $\sum_{i=1}^{6} n_i = n$. The probability that a given combination of the n_i's results is obviously related to the p_i's.

If the society is completely homogeneous then each voter will have the same linear preference ranking on the candidates, say $n_1 = n$. If the society is completely heterogeneous then the voters are expected to be equally divided among all the possible preference rankings on the candidates, so we expect $n_i = \frac{n}{6}$ for all i. Generally speaking, social homogeneity increases as the preference orders of individuals in the society become more similar to one another.

Various measures of social homogeneity will naturally be related to the p_i's. We shall distinguish between two types of measures of social homogeneity. Order specific measures shall use the p_i's along with the information of the specific linear ranking that each is associated with. Non-order specific measures will use only the p_i's without using the knowledge of the specific linear order that each is attached to. When considering a choice of a measure of social homogeneity, one might intuitively expect to have a tradeoff between precision and simplicity. Since the order specific measures use additional information they are likely to give more precise results, at least when making relative comparisons between sets of p_i's. But the non-order specific measures should generally be easier to calculate.

A number of studies have been conducted to examine the relationship between social homogeneity and the probability that there is a Condorcet winner. For a Condorcet winner to exist for any specific profile, the n_i's must be such that some candidate must be able to defeat all remaining candidates in a series of pairwise elections. For example, with $n_1 = 2$ and $n_3 = 1$ for $n = 3$ we find A is the Condorcet winner since it beats B two to one and it beats C three to zero in pairwise comparisons. It is well known that a Condorcet winner does not always exist. For example, with $n_1 = 1$, $n_4 = 1$ and $n_5 = 1$ for $n = 3$ we find A beats B two to one, B beats C two to one, and C beats A two to one.

The probability that a Condorcet winner exists is obviously related to n and the p_i's. Gehrlein and Fishburn (1976a) have shown that the probability that there is a Condorcet winner for three candidates and odd n is given by $P_n(p)$ where for the vector of p_i's given by p

$$P_n(p) = \sum^{3} \frac{n!}{m_1!m_2!m_3!m_4!} \left\{ (p_4+p_6)^{m_1} p_3^{m_2} p_5^{m_3} (p_1+p_2)^{m_4} + (p_2+p_5)^{m_1} p_1^{m_2} p_6^{m_3} (p_3+p_4)^{m_4} + (p_1+p_3)^{m_1} p_4^{m_2} p_2^{m_3} (p_5+p_6)^{m_4} \right\}.$$

Here, a! is a-factorial, $m_4 = n-m_1-m_2-m_3$ and $\overset{3}{\Sigma}$ is the triple sum over m_1, m_2 and m_3 with

$$0 \leq m_1 \leq (n-1)/2$$
$$0 \leq m_2 \leq (n-1)/2-m_1$$
$$0 \leq m_3 \leq (n-1)/2-m_1$$

A more complicated relation for $P_n(p)$ has been developed by DeMeyer and Plott (1970) and $P_n(p)$ relations have been obtained for special assumptions about n and the p_i's by Garman and Kamien (1968), Niemi and Weisberg (1968), Sevcik (1969), Gehrlein and Fishburn (1976b, 1979) and Fishburn and Gehrlein (1978).

A number of studies have suggested that there should tend to be a positive relationship between measures of social homogeneity and the probability of a Condorcet winner. That is, as societies become more homogeneous the probability of a Condorcet winner should also increase. This relationship has been found to hold up for several different profile specific measures of social homogeneity. Niemi (1969) found this relationship by measuring social homogeneity by the maximum number of voters whose preference orders are jointly single peaked. Fishburn measured the homogeneity by using the Kendall-Smith coefficient of concordance (Kendall and Smith (1939)) and found a positive relationship. Jamison and Luce (1972) and Kuga and Nagatani (1974) also found this positive relationship to hold up.

In the current study we concern ourselves with non-profile specific measures of social homogeneity. Several articles have been concerned about the relationship between the probability of a Condorcet winner and the non-profile specific measure of social homogeneity $S^1(p)$ where

$$S^1(p) = \sum_{i=1}^{6} p_i^2 .$$

$S^1(p)$ is maximized with one of the p_i's equal to one with the rest equal to zero, which is complete homogeneity. $S^1(p)$ is minimized by $p_i = 1/6$ for all i which reflects a totally heterogeneous situation. The measure $S^1(p)$ was suggested by Abrams (1976) who showed that a perfect positive relationship does not exist. That is, we can define two sets of p_i's, given by p and p' where $P_n(p)$ is greater than $P_n(p')$ while $S^1(p)$ is less than $S^1(p')$. Fishburn and Gehrlein (1978) consider the $S^1(p)$ vs $P_n(p)$ relation more generally. While in fact specific examples can be developed to show the behavior described by Abrams we might not generally expect this to happen. This study (Fishburn and Gehrlein 1978) was conducted in two steps. The first step was to consider the set of p vectors for which a direct positive $S^1(p)$ and $P_n(p)$ relation holds up. The second step was to look at an indirect relation between $P_3(p)$ and $S^1(p)$ over all p vectors.

When considering the p vectors for which a direct positive $S^1(p)$ and $P_n(p)$ relation exists, the restriction was made that n was infinite. The set of p vectors considered was then limited to those meeting the dual culture condition. If a p vector meets the dual culture condition it must be true that the probability attached to any linear preference order is the same as the probability attached to the dual (all preferences reversed) of that preference order. For our three alternative case we require $p_1 = p_6$, $p_2 = p_4$ and $p_3 = p_5$. It is shown that if $S^1(p)$ is increased by changing two of p_1, p_2 and p_3 while keeping the other fixed then $P_\infty(p)$ also increases if p_4, p_5 and p_6 change accordingly to stay in the space of dual culture vectors.

The indirect relationship between $P_3(p)$ and $S^1(p)$ was done by considering a different measure for $P_3(p)$. The measure used was $Q_3(p)$ which was obtained as the sum of the $P_3(p)$ values where the sum was over the 6! permutations of the p vector. The upper and lower bounds on $Q_3(p)$ were found for fixed $S^1(p)$ and it was found that both of these bounds increase as $S^1(p)$ increases. This suggests a generally positive relation between $Q_3(P)$ and $S^1(p)$.

The purpose of the current study is to examine the general relationship between $P_n(p)$ and $S^1(p)$ over all p vectors. This relationship has only been shown to hold under specific restrictions in the previously mentioned studies.

In addition, other non-profile specific measures of social homogeneity will be developed. These measures will then be compared on the basis of how they relate to $P_n(p)$.

Measures of Social Homogeneity

$S^1(p)$ is the non-profile specific measure of social homogeneity that has received the most attention in the literature. There are many more non-profile specific measures of social homogeneity that might be considered. The measures considered in this study are defined on a probability vector p by

$$S^2(p) = \sum_{i=1}^{6} p_i^3$$

$$S^3(p) = \sum_{i=1}^{6} p_i^4$$

$$S^4(p) = \mathrm{Max}_i \{p_i\}$$

$$S^5(p) = \mathrm{Min}_i \{p_i\}$$

$$S^6(p) = S^3(p) - S^4(p)$$

$$S^7(p) = \prod_{i=1}^{6} p_i$$

$$S^8(p) = \sum_{i=1}^{6} |p_i - \tfrac{1}{6}|$$

By these definitions we see that $S^2(p)$ is the sum of the cubes of the p_i's. This measure is a natural extension of $S^2(p)$. Similarly, $S^3(p)$ is the sum of the fourth powers of the p_i's. $S^4(p)$ is simply the maximum p_i component in p. $S^5(p)$ is the minimum p_i component in p. $S^6(p)$ is the range of the p_i components. $S^7(p)$ is the product of the six p_i components. $S^8(p)$ is the sum of this absolute deviations of the p_i components from their mean, where $\frac{1}{6}$ is the mean p_i for the three candidate case. All eight non-profile specific measures of social homogeneity are fairly common measures of dispersion or variation among a set of numbers. The most common measure of variation in a set of numbers is variance and this is not considered since it is directly related to $S^1(p)$.

As with $S^1(p)$, all of these measures, except $S^5(p)$ and $S^7(p)$, are expected to increase as social homogeneity increases. Both $S^5(p)$ and $S^7(p)$ should tend to decrease as homogeneity increases. In order to determine whether or not a positive relationship exists between the social homogeneity and $P_n(p)$, simulation analysis was employed.

Simulation Format

In order to begin our analysis it is first necessary to generate p vectors. This was done by generating a random number on the interval [0,1) for each of the six p_i's. Since we require $\sum_{i=1}^{6} p_i = 1$, each p_i was normalized by dividing by the sum of the p_i's. Once a p vector was obtained $S^i(p)$ was obtained for i = 1, 2, ..., 8 and $P_n(p)$ was calculated. For each n value of 3, 5, 7, ..., 13; this process was repeated 2500 times. Having these results, the $S^i(p)$ vs $P_n(p)$ relationship can be considered.

Results

The attempt to find a general relationship between $S^i(p)$ and $P_n(p)$ takes place in two steps. In the first step, a relation is considered by using the coefficient of correlation between $P_n(p)$ and each of the $S^i(p)$ measures. The results are presented in Table 1.

Table 1 about here

The results of Table 1 give almost no support to the statement that a positive relationship exists between social homogeneity and $P_n(p)$ for any i. The existence of negative correlations for $S^5(p)$ and $S^7(p)$ exist since, by their definition, increases in these two measures suggest a decrease in homogeneity. Trends in the data suggest that the weak relationship that does exist tends to decrease for all $S^i(p)$ as n increases.

These results could come from two sources. Either there simply is no relationship between $S^i(p)$ and $P_n(p)$ or if there is a relationship it is a non-linear relationship. To test the notion that there is a stronger relation than that suggested by correlation calculations, another measurement can be used. In this stage the 2500 $S^i(p)$ and $P_n(p)$ values were checked to see the proportion of times that each changed in the same direction. That is, start with the first and second p vectors, namely p^1 and p^2. If $S^i(p^1) > S^i(p^2)$ and $P_n(p^1) > P_n(p^2)$ then each changed in the same direction. The same is true if $S^i(p^1) < S^i(p^2)$ and $P_n(p^1) < P_n(p^2)$. By going through all consecutive p vectors in the 2500 observations, we can find the proportion of times that $P_n(p)$ and $S^i(p)$ changed in the same direction. If this proportion is significantly different than .5, a general relationship can be assumed. The results of the second stage of this analysis are presented in Table 2.

Table 2 about here

The results of Table 2 give little support to the statement that a strong positive relationship exists between social homogeneity and $P_n(p)$. As with the correlation results, trends in the data suggest that the weak relationship that does exist tends to decrease for all $S^i(p)$ as n increases. Due to the large sample size, a hypothesis test that any of these proportions differ significantly from .5 would be accepted, even for minute probabilities of a type one error. However, the evidence suggesting a strong positive relationship is very weak.

Conclusion

The results obtained from the simulation analysis suggest that any general relationship between social homogeneity and the probability of a Condorcet winner is very weak. This statement holds for the non-profile specific measures of social homogeneity used in this study. The results of May (1971) suggest that this should be expected for very large numbers of voters. From this study, we see that it is also true for a very small number of voters. An additional point is that none of the eight non-profile specific measures of social homogeneity shows any dominance over the other measures. The strength of the relationship that does exist generally decreases as the number of voters increases. If a strong relationship is to be found between social homogeneity and the probability of a Condorcet winner, it appears that a fairly sophisticated profile specific measure of social homogeneity will be required.

References

1. Abrams, R., The voter's paradox and the homogeneity of individual preference orders, Public Choice, 26 (1976), 19-27.

2. DeMeyer, F. and C. R. Plott, The probability of a cyclical majority, Econometrica, 38 (1970), 345-354.

3. Fishburn, P. C., Voter concordance, simple majority, and group decision methods, Behav. Sci., 18 (1973), 364-376.

4. Fishburn, P. C. and W. V. Gehrlein, Social homogeneity and Condorcet's paradox, Public Choice, (1978).

5. Garman, M. and M. Kamien, The paradox of voting: probability calculations, Behav. Sci., 13 (1968), 306-316.

6. Gehrlein, W. V. and P. C. Fishburn, The probability of the paradox of voting: a computable solution, J. Econ. Theory, 13 (1976a), 14-25.

7. Gehrlein, W. V. and P. C. Fishburn, Condorcet's paradox and anonymous preference profiles, Public Choice, 26 (1976b), 1-18.

8. Gehrlein, W. V. and P. C. Fishburn, Proportions of profiles with a majority candidate, Comp. and Maths. with Appls., 5 (1979), 117-124.

9. Jamison, D. and E. Luce, Social homogeneity and the problem of intransitive majority rule, J. Econ. Theory, 5 (1972), 79-87.

10. Kendall, M. G. and B. B. Smith, The problem of m rankings, Ann. Math. Statist., 10 (1939), 275-287.

11. Kuga, K. and H. Nagatani, Voter antagonism and the paradox of voting, Econometrica, 42 (1974), 1045-1067.

12. May, R. M., Some mathematical results on the paradox of voting, Behav. Sci., 16 (1971), 143-151.

13. Niemi, R. G., Majority decision making with partial unidimensionality, Amer. Pol. Sci. Rev., 63 (1969), 488-497.

14. Niemi, R. G. and H. F. Weisberg, A mathematical solution to the probability of the paradox of voting, Behav. Sci., 13 (1968), 317-323.

15. Sevick, K. E., Exact probabilities of the voter's paradox through seven issues and seven judges, U. of Chicago, Institute for Computer Research Quarterly Report 22 (1969), Section III-B.

Table 1

Correlation Between Measures of Homogeneity
and Probability of a Condorcet Winner

Measure	Voters 3	5	7	9	11	13
$s^1(p)$	.322	.180	.122	.125	.089	.105
$s^2(p)$	.305	.176	.121	.136	.099	.114
$s^3(p)$	.278	.155	.114	.133	.101	.115
$s^4(p)$	.284	.185	.114	.131	.091	.112
$s^5(p)$	-.215	-.103	-.058	-.030	-.003	-.028
$s^6(p)$	.299	.181	.109	.112	.069	.095
$s^7(p)$	-.270	-.133	-.092	-.052	-.026	-.043
$s^8(p)$	.319	.162	.123	.099	.074	.086

Table 2

Proportion of Times That Measures of Homogeneity
and Probability of a Condorcet Winner Change in Same Direction

Measure	Voters 3	5	7	9	11	13
$s^1(p)$	.675	.650	.635	.621	.602	.611
$s^2(p)$	.679	.643	.624	.621	.599	.607
$s^3(p)$	.670	.640	.623	.621	.591	.611
$s^4(p)$	.650	.630	.615	.601	.582	.602
$s^5(p)$	.375	.410	.404	.421	.424	.430
$s^6(p)$	.668	.635	.619	.621	.600	.594
$s^7(p)$	.335	.378	.386	.393	.401	.402
$s^8(p)$	.673	.644	.634	.627	.599	.619