Fractionalization and inter-group differences*

Jo Thori Lind[†]

August 12, 2005

Abstract

Fractionalization has been shown to have a detrimental effect on growth, public goods provision, and redistribution. However, the conventional measure of fractionalization, the Herfindahl index, implicitly assumes that all groups are equally distant. In this paper, I argue that a more appropriate measure of fractionalization should take into account that some groups are more different than others. I present a simple method to estimate these distances from opinion survey data and show an application to the US.

Keywords: Political economy, fractionalization, race, opinion surveys

JEL Classification: D31, D72, E62, H11, J15, Z13

*I thank Kira Börner, Bård Harstad, Simon Lörcher, Kalle Moene, Debraj Ray, Tore Schweder, seminar participants at PRIO and the University of Oslo and the European Public Choice Society annual meeting for comments, suggestions, and helpful discussions.

[†]Department of Economics, University of Oslo, PB 1095 Blindern, 0317 Oslo, Norway. E-mail: j.t.lind@econ.uio.no

1 Introduction

Fractionalization has caught a lot of attention among economists and other social scientists in recent years. The concept is usually defined as the probability that two randomly chosen persons belong to different groups, be it ethnic, religious, linguistic, or other groups. Studies have shown that fractionalization leads to more corruption (Mauro 1995), low growth and bad policies in general (Easterly and Levine (1997), low provision of public goods (Alesina, et al. 1999), less redistribution (Alesina et al. 2001, Lind 2005), less social mixing and activity (Alesina and La Ferrara 2000), lower voluntary contributions to schools (Miguel 2003, Miguel and Gugerty 2003), and higher prevalence of civil war (Elbadawi and Sambanis 2002, Reynal-Querol 2002).¹

To study the effect of fractionalization, it is crucial to measure it properly. Most existing measures of fractionalization are only partially successful in this respect. My objective is to develop a method to construct better measures of fractionalization.

The first step to measure fractionalization is to choose the partitioning of the population into groups. The majority of studies use data on ethnic and linguistic groups collected by Soviet anthropologists in the 1960s reported in the Atlas Narodov Mira, and compiled into a measure of ethno-linguistic fractionalization usually referred to as ELF. Posner (2004), among others, have criticized the groups used by the Atlas. He argues that some groups that are actually the same have been grouped as different whereas groups that are arguably distant are grouped together. Alesina et al. (2003) and Fearon (2003) have recently compiled broader data sets with data of higher quality, at least for the purposes of studies of the consequences of fractionalization. But as Fearon (2003: 197) states, “It rapidly becomes

¹A more detailed survey of this literature can be found in Alesina and La Ferrara (2004)

clear that one must make all manner of borderline-arbitrary decisions, and that in many cases there simply is no right answer to the question: “What are the ethnic groups in this country?””, so a correct way of calculating fractionalization has not yet been found.

For most societies, there are some group partitions that matter for politics, and a large number of other possible partitions that have no relevance for politics. Also, for a given partition, it may be that the division between some groups is more important than the division between others. One way to put the problem is to say that the distance between groups is not necessarily the same between all groups. If we have a concept of distance between groups, this may also help us tell what groups we should use in a proper analysis of fractionalization. If the distance between two groups is small, the splitting into two groups may not be relevant. If it is large, the partitioning is relevant.

The aim of the present paper is to make a first attempt at estimating the distance between groups using data from opinion surveys. The method I suggest is based on estimated differences in opinions on a set of political questions. We regress a measure of political opinions, such as whether expenditure on some public good should be increased or not, on dummies for belonging to the groups of a potentially salient partition of society as well as a vector of control variables. If members of different groups have significantly different opinions, controlling for other characteristics, this indicates that groups in general have different opinions on this dimension of politics, and hence that they are different. Comparing the magnitude of the coefficient among groups, we can construct a measure of distance between groups.

There are usually several relevant opinion questions we wish to include, and each question will in general give a separate measure of inter-group distances. To aggregate these coefficients, I suggest to locate groups in a space with dimensionality equal to the number of

question and the coordinates being given by the estimated coefficients. Then the aggregate distance between groups can be calculated as the distance between points in this space.

The appropriate choice of which questions to study depends on the final objective of the study. For a study of the effect of fractionalization on public goods provision question relating to public policy are of crucial interest. If the objective is the effect of fractionalization on the prevalence of civil war, such questions may be less relevant.

The paper relates to the body of literature studying the consequences of fractionalization on different outcomes mentioned above. On the theoretical level, it relates closely to the work of Caselli and Coleman (2002). They present a model of coalition formation where the success of the formation depends on being able to exclude others. They argue that this is most likely if the coalition is formed by one ethnic groups that is more distant from other ethnic groups. As strong coalition leads to more rent seeking activity, larger distance between groups is economically detrimental.

Recently, a literature attempting to construct better measures of fractionalization has also emerged. One improvement is the updating and increased coverage provided by Alesina et al. (2003) and Fearon (2003). Another is in the measurement of ethnicity and fractionalization.

The fundamental task of placing individuals in ethnic groups, which is required for calculating ethnic fractionalization, is in itself not trivial. As pointed out by Posner (2000), a person's self classification depends on his current context. He suggests a methodology where he creates a uniform context by means of a recorded dialogue before asking respondents about their ethnicity. Nopo et al. (2004) measure respondents intensity along several racial characteristics by physical characteristics. Bannon et al. (2004) use self reported data on which group people feel they belong to, and study who report that their belonging is to their ethnic group.

Posner (2004) argues that the grouping of ethnicities found in the standard sources of data is inappropriate. Some groups are strictly speaking different groups, but cooperate well so they should not be counted as separate. Going through a large amount of literature on each country, he constructs a measure of relevant ethnic groups for a number of African countries. The problem with this approach is mostly the large workload and the difficulty of replicating the data construction. The methodology I derive in this paper is also suitable to identify which groups are very close and which are further apart, so it serves as an alternative.

There is a small literature that attempts to estimate the distance between different linguistic groups, starting with Greenberg's (1956) seminal paper. He suggested to calculate the distance between two linguistic groups (actually two languages) as the fraction of a fixed basic vocabulary, taken from a glottochronology or Swadesh list, that are common between the two. As this approach to linguistics has lost most of its popularity, it has not been pursued much further. Fearon and Laitin (Fearon and Laitin 1999, Laitin 2000) has pursued a related direction where they measure the distance between languages by their distance in linguistic trees. If two languages belong to different language families, they obtain the maximum distance and if they are closely related they get a smaller distance. The drawback with this approach for measuring fractionalization is that two groups may have very different languages but still work perfectly well together, and another society may have groups that speak almost the same language, such as Serbian and Croatian, but still be unable to cooperate.

2 An improved measure of fractionalization

The conventional measure of fractionalization is the Herfindahl measure

$$H = 1 - \sum_{i=1}^N q_i^2, \quad (1)$$

where there are N groups and q_i is the fraction of the population belonging to group i . This measure gives the probability that two randomly chosen individuals belong to different groups. Implicit in this measure is that if two persons belong to the same group, they are in some way identical. If they are from two different groups, the difference between the two is the same independently of which groups they belong to.

However, it is often the case that some groups are close and other more remote. Panel a of Figure 1 depicts the case where H may be appropriate. Here, all three groups have the same distance. If we say that each group consists of $1/3$ of the population, total fractionalization is $2/3$. In panel b, groups B and C are closer. Here H would still give a fractionalization of $2/3$. In panel c, however, groups B and C are even closer together and have merged to one group. Now H measures fractionalization as $4/9$. This illustrates two difficulties with the Herfindahl measure H . First, if some groups are close to each other, this should be reflected in the measure of fractionalization, so fractionalization in case b should be lower than in case a. Second, there is a discontinuity in the measure. It is often not clear exactly how many groups we should have. We could have groups that are close. Then fractionalization should not change abruptly if we put them in two different groups, as in panel b, or in one group, as in panel c.

An example of division where the distance is small could be the Norwegian division between bokmål and nynorsk speakers. The two languages are very similar, and speakers of one language can easily understand speakers of the other language. Still both have the

Figure 1: *Different cases of group division.*

status of official languages. With approximately 15% nynorsk users², this gives a Herfindahl measure of 0.26. We can compare this with for instance the Basque minority in Spain³, which only makes up 1.6% (Alesina et al. 2003) and hence yields a Herfindahl measure of 0.03. Most observers would agree that the linguistic fractionalization in Norway has virtually no effect whereas the fractionalization in Spain is a salient conflict. The reason is of course that the distance between the two linguistic groups in Norway is small whereas the distance between the groups in Spain is much larger.

An extension of (1) that can solve these problems is

$$F = \sum_{i=1}^N \sum_{j=1}^N d_{ij} q_i q_j, \quad (2)$$

where d_{ij} is a measure of the distance between group i and j . It is natural to assume that $d_{ii} = 0$, i.e. that each group is homogeneous. Also, we usually have symmetric distances, so $d_{ij} = d_{ji}$. In Figure 1, panel a) would correspond to the case where all the distances

²To the best of my knowledge, there are no exact measures of the population using nynorsk, partly because some use both. The number is based on the fraction of pupils using nynorsk (Statistics Norway 2004).

³One could argue that other Spanish minorities, e.g. the Catalans, should be considered separate groups. However it seems that the conflict between the Basques and the remainder is more salient.

d_{ij} , $i, j \in \{A, B, C\}$ are unity, so fractionalization is $2/3$ as with the Herfindahl measure.

In panel b, however, we still have $d_{AB} = d_{AC} = 1$, but now $d_{BC} < 1$ so fractionalization is below $2/3$ as measured by the measure F . As the distance d_{BC} shrinks, we converge to the case depicted in panel c, where $d_{BC} = 0$ in a logical way.

3 Relation to existing measures

It is easily seen that the new fractionalization measure F reduces to the Herfindahl measure H if we choose the distance function

$$d_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases}, \quad (3)$$

which implies that the distance is the same between all groups.

It has been suggested that some features are better explained by polarization than fractionalization. The measure F can be extended to encompass the class of polarization measures introduced by Esteban and Ray (1994). If we define

$$F^\alpha = \sum_{i=1}^N \sum_{j=1}^N d_{(k_j, l_i)} i j q_i^{1+\alpha} q_j,$$

we would get their measure of polarization⁴ for $\alpha \in (0, \alpha^*]$ and the measure of distance polarization with $\alpha = 0$. If we set $\alpha = 1.5$ and restrict d to be the binary metric (3), which is also used for the Herfindahl index, we get the measure of group polarization suggested by Reynal-Querol (2002). I conjecture that one way to derive this measure theoretically could be through Esteban and Ray's (1999) model of the effect of polarization on conflict.

⁴Where α^* is some constant $\simeq 1.6$.

4 Measuring inter-group distance

The measure of fractionalization F introduced in (2) is theoretically appealing, but not applicable for practical purposes unless we find a way to determine the distances d_{ij} between groups. One could think of several ways to measure distances between groups, and it is likely that different measures would be appropriate for different purposes.

One approach that is widely applicable is to use stated preferences on policy questions. If members of different groups have very different opinions on these questions, holding other characteristics constant, this indicates that there is a large distance between these groups. Of course, it can be the case that they have very different opinions on some aspects of politics and more similar views on other aspects. This would then mean that they have a large distance along some dimensions, but are closer along other dimensions. For instance, one measure of inter group distances may be appropriate for studying provision of public goods, but not for the probability of violent conflict.

Here I show how we can use opinion survey data to measure the distance between a given group and a reference group. If we regress respondents' opinion on some political question, such as whether public goods provision is too high or too low, on a vector of standard control variables and dummy variables for the respondent's group, the coefficient on the group dummies can be interpreted as the distance between the group and the reference group (the omitted category).

If the estimated parameters for two groups are very similar, so their distance is estimated to be small, this shows that these groups tend to have similar opinions on this question and hence that a coalition of these two groups is likely. Which potential coalition formation we are interested in depends on the ultimate goal of our analysis, and determines the appropriate

choice of questions to analyse.

If Y_i is an indicator of opinions on some policy issue, g_{ij} an indicator of individual i belonging to group j and z_i a vector of control variables, we run a regression of the form

$$Y_i = \alpha + \sum_j \delta_j q_{ij} + \gamma' z_i + \varepsilon_i. \quad (4)$$

If we let group 0 be the reference group and impose $\delta_0 = 0$, we can define the distance between group j and group k as $|\delta_j - \delta_k|$.

A difficulty with this approach is that there will usually be several questions on opinion that are useful to characterize distances between groups and a lot of information will disappear if we only choose to use one indicator. This means we actually want to estimate a system of the form

$$\begin{cases} Y_{i1} = \alpha_1 + \sum_j \delta_{j1} q_{ij} + \gamma'_1 z_i + \varepsilon_{i1} \\ \vdots \\ Y_{iT} = \alpha_T + \sum_j \delta_{jT} q_{ij} + \gamma'_T z_i + \varepsilon_{iT} \end{cases},$$

i.e. one equation for each of T opinion indicators. For each group j we then have T distance indicators δ_{jt} , and we usually want to aggregate these into a single distance measure. A simple way to do this is to let each of the T indicators represent a dimension in Euclidean space, so $(\delta_{j1}, \dots, \delta_{jT})$ is group j 's location. The case of three groups and two indicators is shown in Figure 2. Here group 0 and 1 are clearly the furthest apart. Group 2 is closest to group 1 along dimension 1 and closest to group 0 along dimension 2. The overall distances between the groups d_{jk} is determined as the distance between the points.

Although the situation is more difficult to depict graphically, it is straightforward to analyse the situation with more than two opinion indicators. The aggregate distance measure

Figure 2: *Estimation of inter-group distance.*

is then calculated by the Euclidean metric

$$d_{jk} = \sqrt{\sum_t (\delta_j - \delta_k)^2}. \quad (5)$$

This could be extended to other metrics as well as different weighting schemes, but as it is usually difficult to choose these, I prefer to stick to the simple method.⁵

⁵A different approach would be to first estimate a racial index r_i and then use this in the opinion

5 An application to US data

A rich data set that permits estimating inter-group distances is the US General Social Survey (GSS). I use this to estimate a politically relevant measure of the distance between the groups of African Americans, Whites, and Others. The GSS, which is an annual omnibus-type survey which has been conducted since 1972, contains a number of questions on opinions of public policy that may seem relevant for estimating inter-group distances. Table 1 gives an overview of the opinion questions I use.

Table 1 about here

I regress each of these measures on dummies for race, which is likely to be the most relevant group decomposition in the US case, as well as a set of other variables we believe have an impact on opinions as in equation (4). Results from these analyses are reported in Table 2. To keep matters as simple as possible, I use a linear probability model even

regression. Specifically, we can think of this as a model on the form

$$\begin{aligned}y_{it} &= \alpha_t + \beta_t r_i + \gamma'_t z_i + \varepsilon_{it} \\ r_i &= \sum_j \delta_j q_{ij}\end{aligned}$$

Stacking vectors by question we can rewrite the model as

$$y_i = \alpha + \Phi d_i + \gamma z_i + \varepsilon_i$$

where we have the restriction $\Phi = \beta\delta$, which essentially entails that Φ should have rank 1. This is a problem of reduced rank regression introduced by Anderson (1951); see e.g. Reinsel and Velu (1998) for an updated and more accessible treatment. Estimation is carried out by first using the Frisch-Waugh theorem to partial out the control variables z_i . An unrestricted version of Φ is obtained by OLS, and then δ is obtained as a vector weighted average of the unrestricted Φ . Estimation results from this method is provided in Appendix table 1.

though the respondents give ranges of answers. Details of the coding can be found in Table 1. Generally, I have tried to code an answer that could be seen as liberal as a positive outcome.

Table 2 about here

On all issues except tax policy, African Americans seem to have a more liberal view than Whites as their coefficient is positive, and also significant. This is also the case for the Other group on most issues, but here not all estimates are significantly different from zero. This shows that political opinions tend to be more homogeneous within one racial group than in the population as a whole. This coefficient on the group dummies may be interpreted as the distance between the group and Whites, and the distance between African Americans and Others would be found as the (absolute value of the) difference of the coefficients on the two groups. However, now we have a large number of distance measures, which is awkward in itself. Furthermore, it is not trivial to decide what measure to use for a particular issue.

Table 3 about here

To try to construct a more useful measure of distance, we can first notice that there is a high degree of association between the different questions in the battery to measure opinions. Table 3 show the results from a factor analysis of the answers. We see that there is one dominant factor, which seems to correspond reasonably well to the liberal/conservative distinction. Now, we could extract this factor, and use this as the explained variables rather than the whole battery. However, the control variables have different impacts on different opinion questions. As it is crucial to control for other factors to get an estimate of group membership proper on opinions, this could lead to flawed estimates. Hence the Euclidean metric (5) is calculated from the estimates reported in Table 2. The resulting aggregate

distances are reported in Table 4.

Table 4 about here

These estimates would be useful to construct a measure of fractionalization for the whole of the US. However, to get a better grasp of the performance of the procedure, it is useful to introduce some cross sectional variation. To do this, I assume that the distance parameters δ_{jk} may vary across census regions. Using a similar technique as the one used to produce Table 4 gives the distance measures shown in Figure 1. The three first panels show the distances between the three groups geographically, whereas the last panel show the actual numbers of the distances between African Americans and Whites and Whites and Others.

Now it is time to return to the original task, that of constructing a theoretically more appealing measure of fractionalization. Using population estimates for 2003⁶, it is straightforward to construct the measure of distance fractionalization F using the distance estimates above. Figure 2 shows the geographical distribution of the traditional Herfindahl measure of fractionalization and the new distance measure as well as a plot of the two against each other. We see that the two measures are highly correlated (the correlation coefficient is .90). To get a grasp of the differences between the two measures, the last panel shows the geographical distribution of residuals from a regression of distance fractionalization on the Herfindahl measure. If these residuals are positive, it indicates that the Herfindahl index underestimates the fractionalization in this state and vice versa for negative values. We see that the Herfindahl measure tends to overestimate the degree of fractionalization on the West Coast and more markedly in Hawaii and Alaska. It tends to underestimate the level of fractionalization in most of the South. This conforms well with the received wisdom that

⁶Taken from <http://www.census.gov/popest/states/asrh/tables/SC-EST2003-04.xls>

Figure 3: Estimates distances between racial groups.

The maps show distance between racial group estimated as the Euclidean distance using the results from the regressions.

Figure 4: Fractionalization across states

The panel Difference show the residuals from a regression of distance fractionalization on the Herfindahl index. A positive value indicates that the Herfindahl index underestimates fractionalization.

racial conflict is more intense in the South than along the West Coast.

6 Conclusion

In this paper, I argue that the conventional measure of fractionalization, the Herfindahl index, is too simplistic, and suggest a more general measure that gives the average distance between groups. This reduces to the conventional measure if the distance between all groups is identical.

To estimate between group distances, I derive a method based on regressing political opinions on dummies for group membership. If a dummy variable on a given group have a large coefficient, other explanatory factors controlled for, it indicates a large distance between the groups. As there are several opinions we would like to use for such measurement, we need to combine these measures into a single measure of distance. I show how to do this by letting the different political issues correspond to dimensions in an Euclidean space.

I used this approach to construct new measures of fractionalization for US states based on opinion data from the GSS. The new data seem to give a better picture of fractionalization than the conventional measure.

In future work, it would be very useful to construct measures of inter-group distance and distance-fractionalization on a broader sample of countries. The challenger is to get comparable opinion data for an interesting sample of countries. Some fairly uniform multi-country opinion surveys have emerged (World Values Survey, Afrobarometer) and could be potential sources of such data. However, the data coverage for developing countries is still fairly low.

References

- Alesina, A., R. Baqir, and W. Easterly (1999): “Public goods and ethnic division.” *Quarterly Journal of Economics* 114: 1243-84.
- Alesina, A., A. Devleeshauwer, W. Easterly, S. Kurlat, and R. Waciarg (2003): “Fractionalization.” *Journal of Economic Growth* 8: 155-94.
- Alesina, A., E. Glaeser, and B. Sacerdote (2001): “Why doesn’t the US have a European-style welfare system.” *Brookings Papers on Economic Activity* 187-278.
- Alesina, A., and E. La Ferrara (2000): “Participation in Heterogeneous Communities.” *Quarterly Journal of Economics* 115: 847-904.
- Alesina, A. and E. La Ferrara (2004): “Ethnic diversity and economic performance.” Mimeo, Harvard University.
- Anderson, T. W. (1951): “Estimating linear restrictions on regression coefficients for multivariate normal distributions.” *Annals of Mathematical Statistics* 22: 327-51.
- Bannon, A., E. Miguel, and D. N. Posner (2004): “Sources of ethnic identification in Africa.” Mimeo, Berkeley.
- Easterly, W., and R. Levine (1997): “Africa’s Growth Tragedy: Policies and Ethnic Divisions.” *Quarterly Journal of Economics*, 112: 1203–1250.
- Elbadawi, I., and N. Sambanis (2002): “How much civil war will we see?” *Journal of Conflict Resolution* 46: 307-34.
- Esteban, J.-M. and D. Ray (1994): “On the Measurement of Polarization.”

- Econometrica* 62: 819-51.
- Esteban, J.-M. and D. Ray (1999): “Conflict and distribution.” *Journal of Economic Theory* 87: 379-415.
- Fearon, J. D. (2003): “Ethnic and cultural diversity by country.” *Journal of Economic Growth* 8: 195-222.
- Fearon, J. D., and D. D. Laitin (1999): “Weak states, rough terrain, and large-scale ethnic violence since 1945.” Mimeo, Stanford University.
- Greenberg, J. H. (1956): “The measurement of linguistic diversity.” *Language* 32: 109-15.
- Laitin, D. D. (2000): “What is a language community?” *American Journal of Political Science* 44: 142-55.
- Lind, J. T. (2005): “Fractionalization and the size of government.” Mimeo, University of Oslo.
- Mauro, P. (1995): “Corruption and Growth.” *Quarterly Journal of Economics*, 110: 681–712.
- Miguel, E. (2003): “Tribe or Nation? Nation-Building and Public Goods in Kenya versus Tanzania.” Unpublished paper.
- Miguel, E. and M. K. Gugerty (2003): “Ethnic Diversity, Social Sanctions, and Public Goods in Kenya.” Unpublished paper.
- Nopo, H., J. Saavedra, and M. Torero: “Ethnicity and earnings in Peru.” Mimeo, Middlebury College.
- Posner, D. N. (2000): “Measuring ethnic identities regarding inter-group rela-

tions: Methodological pitfalls and a new technique.” Mimeo, UCLA.

Posner, D. N. (2004): “Measuring ethnic fractionalization in Africa.” *American Journal of Political Science* 48: 849-63.

Reinsel, G. C., and R. P. Velu (1998): *Multivariate Reduced-Rank Regression. Theory and Applications*. New York: Springer.

Reynal-Querol, M. (2002): “Ethnicity, Political Systems, and Civil Wars.” *Journal of Conflict Resolution* 46: 29-54

Statistics Norway (2004): “Grunnskoleelevar, etter elevens målform og fylke. 1. oktober 2002”, <http://www.ssb.no/aarbok/tab/t-040220-183.html> (accessed 20.4.05).

Table 1: Questions on political opinions

	Question	Scale	Coded as 1	GSS shorthand
Liberal	Self classification of liberal or conservative	Scale 1-7	<=4	polview
Environment	Spending on Improving and protecting the environment	Too little, about right, too much	Too little	natenvi
Health	Spending on improving and protecting the nation's health	Too little, about right, too much	Too little	natheal
Cities	Spending on solving the problems of the big cities	Too little, about right, too much	Too little	natcity
Crime	Spending on halting the rising crime rate	Too little, about right, too much	Too little	natcrim
Drug	Spending on dealing with drug addiction	Too little, about right, too much	Too little	natdrug
Education	Spending on improving the nation's education system	Too little, about right, too much	Too little	nateduc
Race	Spending on improving the condition of Blacks	Too little, about right, too much	Too little	natrace
Foreign aid	Spending on foreign aid	Too little, about right, too much	Too little	nataid
Welfare	Spending on welfare	Too little, about right, too much	Too little	natfare
Roads	Spending on highways and bridges	Too little, about right, too much	Too little	natroad
Arms	Spending on military, armaments, and defense	Too little, about right, too much	Too much	natarms
Transport	Spending on mass transportation	Too little, about right, too much	Too little	natmass
Parks	Spending on parks and recreation	Too little, about right, too much	Too little	natpark
Social security	Spending on social security	Too little, about right, too much	Too little	natsoc
Inequality	Government should do something to reduce differences between rich and poor	Scale 1-7	<=3	eqwlth
Tax	Federal tax you pay to high or too low	Too high, right, too low	Right, too low	tax

Table 2: The effect of group belonging on political opinions

		African American	Other non- white	Observations	R ²
(1)	Liberal	0.0568 (6.57)**	0.0452 (2.94)**	31886	0.03
(2)	Environment	0.0505 (4.84)**	-0.0343 (1.64)	22337	0.09
(3)	Health	0.1104 (10.70)**	-0.0433 (2.09)*	22679	0.04
(4)	Cities	0.1732 (15.66)**	0.0318 (1.42)	20626	0.06
(5)	Crime	0.0616 (6.12)**	0.0156 (0.78)	22479	0.02
(6)	Drug	0.1083 (10.20)**	0.0524 (2.45)*	22248	0.02
(7)	Education	0.1177 (11.43)**	-0.0424 (2.04)*	22775	0.08
(8)	Race	0.5427 (57.05)**	0.1130 (5.66)**	21649	0.19
(9)	Foreign aid	0.0470 (9.64)**	0.0285 (2.89)**	22359	0.03
(10)	Welfare	0.2187 (26.23)**	0.0324 (1.93)	22490	0.10
(11)	Roads	0.0144 (1.34)	-0.0705 (4.18)**	20996	0.03
(12)	Arms	0.0866 (8.50)**	0.0586 (2.87)**	22208	0.08
(13)	Transport	0.0406 (3.80)**	0.0153 (0.91)	19831	0.05
(14)	Parks	0.1345 (13.06)**	0.0379 (2.36)*	21120	0.04
(15)	Social security	0.1446 (13.67)**	0.0174 (1.03)	20992	0.09
(16)	Inequality	0.1372 (11.87)**	0.0866 (4.50)**	18775	0.06
(17)	Tax	-0.1134 (10.76)**	-0.0299 (1.66)	20630	0.05

Control variables are age, age squared, log income, log income squared, years of education, years of education squared, number of children, and dummies for sex, having no children, residential density, census region, marital status, and year. t-values in parenthesis, * denotes significant at the 5% level and ** significant at the 1% level.

Table 3: Factor analysis of opinion dimensions

	Factor			Uniqueness
	1	2	3	
Liberal	0.10	0.24	-0.01	0.93
Environment	0.66	0.02	-0.01	0.57
Health	0.71	0.02	-0.04	0.49
Cities	0.63	0.01	0.03	0.60
Crime	0.71	-0.10	-0.18	0.46
Drug	0.69	-0.07	-0.14	0.50
Education	0.68	0.06	-0.02	0.53
Race	0.57	0.09	0.20	0.63
Foreign aid	0.19	0.04	0.18	0.93
Welfare	0.40	0.09	0.25	0.76
Roads	-0.07	0.40	-0.09	0.83
Social security	-0.07	0.48	-0.05	0.76
Inequality	-0.01	0.28	0.04	0.92
Tax	0.00	-0.17	0.12	0.96
Eigenvalue	3.32	0.60	0.22	

Table 4: Estimated Euclidian distances

	Distance
African American - White	0.708 (0.013)
Other - White	0.211 (0.020)
African American - Other	0.588 (0.029)

Numbers are estimated distances using the regression coefficients from Table 2 with standard errors based on 100 bootstrap replications in parenthesis.

Appendix table 1: Results from reduced rank regression

	Estimate	Standard error	Z-value
β			
Liberal	8.95	1.38	6.48
Environment	8.86	1.43	6.19
Health	15.37	1.36	11.33
Cities	26.95	1.35	20.02
Crime	10.05	1.38	7.26
Drug	13.39	1.41	9.50
Education	17.74	1.39	12.72
Race	66.42	0.66	100.61
Foreign aid	5.77	0.66	8.81
Welfare	27.38	1.12	24.52
Roads	0.26	1.13	0.23
Social security	17.78	1.23	14.45
Inequality	13.30	1.28	10.42
Tax	-11.78	1.32	-8.90
δ			
African American	0.00518	0.00009	57.79
Others	0.00101	0.00017	5.88

The first panel gives the effect of group distances on opinions on each of the questions in the battery. The second panel gives the estimated differences between groups.

Standard errors are calculated by bootstrapping with 500 replications. Control variables are dummy for female, age, age squared, log income, log income squared, years of education, years of education squared, number of children, and dummies for no children, marital status, region of residence, and year.