POLARIZATION: CONCEPTS, MEASUREMENT, ESTIMATION

BY JEAN-YVES DUCLOS, JOAN ESTEBAN, AND DEBRAJ RAY1

We develop the measurement theory of polarization for the case in which income distributions can be described using density functions. The main theorem uniquely characterizes a class of polarization measures that fits into what we call the “identity-alienation” framework, and simultaneously satisfies a set of axioms. Second, we provide sample estimators of population polarization indices that can be used to compare polarization across time or entities. Distribution-free statistical inference results are also used in order to ensure that the orderings of polarization across entities are not simply due to sampling noise. An illustration of the use of these tools using data from 21 countries shows that polarization and inequality orderings can often differ in practice.

KEYWORDS: Polarization, inequality, conflict, group identity, axiomatics.

1. INTRODUCTION

INITIATED BY ESTEBAN AND RAY (1991, 1994), Foster and Wolfson (1992), and Wolfson (1994), there has been a recent upsurge of interest in the measurement of polarization2 and in the use of such measures as a correlate of different aspects of socioeconomic performance. It seems fairly widely accepted that polarization is a concept that is distinct from inequality, and that—at least in principle—it could be connected with several aspects of social, economic, and political change.3

Following Esteban and Ray (1991, 1994), we rely almost exclusively on what might be called the identification-alienation framework. The idea is simple: polarization is related to the alienation that individuals and groups feel from one another, but such alienation is fuelled by notions of within-group identity. In concentrating on such phenomena, we do not mean to suggest that instances in which a single isolated individual runs amok with a machine gun are rare, or

1This research was funded by the Pew Charitable Trusts, CRSH, FQRSC, the Chair of Canada in Social Policies and Human Resources, and National Science Foundation 0241070 (Ray). Duclos and Ray thank the Instituto de Análisis Económico (CSIC) for hospitality during the startup phase of this project. Esteban is a member of Barcelona Economics and thanks the support from the Generalitat de Catalunya, the European Commission, the Instituto de Estudios Fiscales, and the MCYT. We thank Oliver Linton, Patrick Richard, a coeditor, and two anonymous referees for useful comments. Finally, we are grateful to Nicolas Beaulieu for his excellent research assistance.


that they are unimportant in the larger scheme of things. It is just that these are not the objects of our enquiry. We are interested in the correlates of organized, large-scale social unrest—strikes, demonstrations, processions, widespread violence, and revolt or rebellion. Such phenomena thrive on differences, to be sure. But they cannot exist without notions of group identity either.

This brief discussion immediately suggests that inequality, inasmuch as it concerns itself with interpersonal alienation, captures but one aspect of polarization. To be sure, there are some obvious changes that would be branded as both inequality- and polarization-enhancing. For instance, if two income groups are further separated by increasing economic distance, inequality and polarization would presumably both increase. However, local equalizations of income differences at two different ranges of the income distribution will most likely lead to two better-defined groups—each with a clearer sense of itself and the other. In this case, inequality will have come down but polarization may be on the rise.

The purpose of this paper is two-fold. First, we develop the measurement theory of polarization for the case in which the relevant distributions can be described by density functions. There are many such instances, the most important being income, consumption, and wealth—regrouped under “income” for short. The reason for doing so is simple: with sample data aggregated along income intervals, it is unclear how to provide a statistically satisfactory account of whether distributive measures (based on such data) are significantly different across time or entities. Indeed, a rapidly burgeoning literature on the statistics of inequality and poverty measurement shows how to construct appropriate statistical tests for such measures using disaggregated data (see, e.g., Beach and Davidson (1983), Beach and Richmond (1985), Bishop, Chakraborti, and Thistle (1989), Kakwani (1993), Anderson (1996), and Davidson and Duclos (1997, 2000)). A rigorous axiomatic development of the polarization concept in the “density case” is then a prerequisite for proper statistical examination of polarization.

In this paper we concentrate on the axiomatics and estimation of “pure income polarization,” that is, of indices of polarization for which individuals identify themselves only with those with similar income levels. (However, Section 4 does contain several preliminary remarks on the broader concept of “social polarization.”) With this settled, we turn to issues of estimation. The main problem is how to estimate the size of the groups to which individuals belong. Again, using arbitrary income intervals would appear somewhat unsatisfactory. Instead, we estimate group size nonparametrically using kernel density procedures. A natural estimator of the polarization indices is then given by substituting the distribution function by the empirical distribution function. Assuming that we are using a random sample of independently and identically distributed observations of income, the resulting estimator has a limiting normal distribution with parameters that can be estimated free of assumptions on
the true (but unknown) distribution of incomes. Distribution-free statistical inference can then be applied to ensure that the orderings of polarization across entities are not simply due to sampling noise.

It is useful to locate this paper in the context of the earlier step in the measurement of polarization in Esteban and Ray (1994)—ER from now on. The measure derived in ER was based on a discrete, finite set of income groupings located in a continuous ambient space of possible income values. This generated two major problems, one conceptual and the other practical. At the conceptual level we have the drawback that the measure presents an unpleasant discontinuity. This is precisely due to the fact that ER is based on a population distributed over a discrete and distinct number of points. The practical difficulty is that the population is assumed to have already been bunched in the relevant groups. This feature rendered the measure of little use for many interesting problems. As mentioned above, the present paper addresses both problems and provides what we hope is a useable measure.

In addition, the main axioms that we use to characterize income polarization are substantially different from ER (though they are similar in spirit). In large part, this is due to the fact that we are dealing with a completely different domain (spaces of densities). We therefore find it of interest that these new axioms end up characterizing a measure of polarization that turns out to be the natural extension of ER to the case of continuous distributions. At a deeper level, there are, however, important differences, such as the different bounds on the “polarization-sensitivity” parameter $\alpha$ that are obtained.

In Section 2 we axiomatically characterize a measure of pure income polarization and examine its properties; this is the conceptual heart of the paper. We then turn in Section 3 to estimation and inference issues for polarization measures and subsequently illustrate the axiomatic and statistical results using data drawn from the Luxembourg Income Study (LIS) data sets for 21 countries. We compute the Gini coefficient and the polarization measure for these countries for years in Wave 3 (1989–1992) and Wave 4 (1994–1997), and find inter alia that the two indices furnish distinct information on the shape of the distributions. Section 4 summarizes the results and discusses an important extension. All proofs are relegated to the Appendix.

2. MEASURING INCOME POLARIZATION

The purpose of this section is to proceed towards a full axiomatization of income polarization.

\footnote{ER (Section 4, p. 846) mention this problem.}

\footnote{In Esteban, Gradin, and Ray (1998) we presented a statistically reasonable way to bunch the population in groups and thus make the ER measure operational. Yet, the number of groups had to be taken as exogenous and the procedure altogether had no clear efficiency properties.}
2.1. Starting Point

The domain under consideration is the class of all continuous (unnormalized) densities in \( \mathbb{R}_+ \), with their integrals corresponding to various population sizes. Let \( f \) be such a density; we are interested in its polarization \( P(f) \). We first describe the notions of “alienation” and “identification” for each individual with income located in the support of \( f \).

We presume that an individual located at \( x \) feels alienation vis-a-vis another located at \( y \), and that this alienation is monotonic in distance \( |x - y| \). This notion is commonplace in the literature on the conceptual foundations of inequality (see, e.g., Sen (1997)).

At the same time, the “identity-alienation framework” we adopt (referred to as IA henceforth) emphasizes that alienation per se is not the end of the story: for alienation to be translated into effective voice, action, or protest, the individual must—to greater or lesser degree—identify with others in society. In this paper, we presume that an individual located at income \( x \) experiences a sense of identification that depends on the density at \( x \), \( f(x) \).

Taken in a broader context, the identification assumption is obviously quite specific. For instance, one might consider the possibility that individuals have a nondegenerate “window of identification” (though the foundations for the width of such an identification window appear unclear). We address this issue (and others) in our discussion of identification in Section 2.4, but recognize that a full analysis of the behavioral foundations of identification is beyond the scope of this paper.

As in ER, we are interested in the effective antagonism of \( x \) towards \( y \) (under \( f \)). In its most abstract form, we may depict this as some nonnegative function

\[
T(i, a),
\]

where \( i = f(x) \) and \( a = |x - y| \). It is assumed that \( T \) is increasing in its second argument and that \( T(i, 0) = T(0, a) = 0 \), just as in ER. (This last condition asserts that while the consequences of an isolated individual’s sense of alienation might be important, this is not the focus of our exercise.) We take polarization to be proportional to the “sum” of all effective antagonisms:

\[
P(F) = \int \int T(f(x), |x - y|) f(x) f(y) \, dx \, dy.
\]

This class of measures is neither very useful nor operational, though at this stage it incorporates the structure of the IA assumptions. In particular, much depends on the choice of the functional form \( T \). In what follows, we place axioms on this starting point so as to pin down this functional form.
2.2. Axioms

Densities and Basic Operations

Our axioms will largely be based on domains that are unions of one or more very simple densities $f$ that we will call basic densities. These are unnormalized (by population), are symmetric and unimodal, and have compact support.\(^6\)

To be sure, $f$ can be population rescaled to any population $p$ by simply multiplying $f$ pointwise by $p$ to arrive at a new distribution $pf$ (unnormalized). Likewise, $f$ can undergo a slide. A slide to the right by $x$ is just a new density $g$ such that $g(y) = f(y - x)$. Likewise for a slide to the left. And $f$ with mean $\mu'$ can be income rescaled to any new mean $\mu$ that we please as follows: $g(x) = (\mu'/\mu)f(x\mu'/\mu)$ for all $x$.\(^7\) These operations maintain symmetry and unimodality and therefore keep us within the class of basic densities.

If we think of slides and scalings as inducing a partition of the basic densities, each collection of basic densities in the same element of the partition may be associated with a root, a basic density with mean 1 and support $[0, 2]$, with population size set to unity. That is, one can transform any basic density to its root by a set of scalings and slides. (This concept will be important both in the axioms as well as in the main proof.) Two distinct roots differ in “shape,” a quality that cannot be transformed by the above operations.

Finally, we shall also use the concept of a squeeze, defined as follows. Let $f$ be any basic density with mean $\mu$ and let $\lambda$ lie in $(0, 1]$. A $\lambda$-squeeze of $f$ is a transformation as follows:

$$f^\lambda(x) \equiv \frac{1}{\lambda} f\left(\frac{x - [1 - \lambda]\mu}{\lambda}\right).$$

A $(\lambda)$-squeeze is, in words, a very special sort of mean-preserving reduction in the spread of $f$. It concentrates more weight on the global mean of the distribution, as opposed to what would be achieved, say, with a progressive Dalton transfer on the same side of the mean. Thus a squeeze truly collapses a density inwards towards its global mean. The following properties can be formally established: (a) For each $\lambda \in (0, 1)$, $f^\lambda$ is a density; (b) for each $\lambda \in (0, 1)$, $f^\lambda$ has the same mean as $f$; (c) if $0 < \lambda < \lambda' < 1$, then $f^\lambda$ second-order stochastically dominates $f^{\lambda'}$; and (d) as $\lambda \downarrow 0$, $f^\lambda$ converges weakly to the degenerate measure granting all weight to $\mu$.

Notice that there is nothing in the definition that requires a squeeze to be applied to symmetric unimodal densities with compact support. In principle, a squeeze as defined could be applied to any density. However, the axioms to be placed below acquire additional cogency when limited to such densities.

---

\(^6\)By symmetry we mean that $f(m - x) = f(m + x)$ for all $x \in [0, m]$, where $m$ is the mean and by unimodality we mean that $f$ is nondecreasing on $[0, m]$.

\(^7\)The reason for this particular formulation is best seen by examining the corresponding cumulative distribution functions, which must satisfy the property that $G(x) = F(x\mu'/\mu)$, and then taking derivatives.
Statement of the Axioms

We will impose four axioms on the polarization measure.

AXIOM 1: If a distribution is composed of a single basic density, then a squeeze of that density cannot increase polarization.

Axiom 1 is self-evident. A squeeze, as defined here, corresponds to a global compression of any basic density. If only one of these makes up the distribution (see Figure 1), then the distribution is globally compressed and we must associate this with no higher polarization. Viewed in the context of our background model, however, it is clear that Axiom 1 is going to generate some interesting restrictions. This is because a squeeze creates a reduction in inter-individual alienation but also serves to raise identification for a positive measure of agents—those located “centrally” in the distribution. The implied restriction is, then, that the latter’s positive impact on polarization must be counterbalanced by the former’s negative impact.

Our next axiom considers an initial situation (see Figure 2) composed of three disjoint densities all sharing the same root. The situation is completely symmetric, with densities 1 and 3 having the same total population and with density 2 exactly midway between densities 1 and 3.

AXIOM 2: If a symmetric distribution is composed of three basic densities with the same root and mutually disjoint supports, then a symmetric squeeze of the side densities cannot reduce polarization.

In some sense, this is the defining axiom of polarization, and may be used to motivate the concept. Notice that this axiom argues that a particular “local”
squeeze (as opposed to the “global” squeeze of the entire distribution in Axiom 1) must not bring down polarization. At this stage there is an explicit departure from inequality measurement.

Our third axiom considers a symmetric distribution composed of four basic densities, once again all sharing the same root.

**Axiom 3:** Consider a symmetric distribution composed of four basic densities with the same root and mutually disjoint supports, as in Figure 3. Slide the two middle densities to the side as shown (keeping all supports disjoint). Then polarization must go up.

Our final axiom is a simple population-invariance principle. It states that if one situation exhibits greater polarization than another, it must continue to do so when populations in both situations are scaled up or down by the same amount, leaving all (relative) distributions unchanged.

**Axiom 4:** If \( P(F) \geq P(G) \) and \( p > 0 \), then \( P(pF) \geq P(pG) \), where \( pF \) and \( pG \) represent (identical) population scalings of \( F \) and \( G \), respectively.
2.3. Characterization Theorem

THEOREM 1: A measure $P$, as described in (1), satisfies Axioms 1–4 if and only if it is proportional to

$$
(3) \quad P_\alpha(f) \equiv \int \int f(x)^{1+\alpha} f(y)|y - x| \, dy \, dx,
$$

where $\alpha \in [.25, 1]$.

2.4. Discussion

Several aspects of this theorem require extended discussion.

Scaling

Theorem 1 states that a measure of polarization satisfying the preceding four axioms has to be proportional to the measure we have characterized. We may wish to exploit this degree of freedom to make the polarization measure scale-free. Homogeneity of degree zero can be achieved, if desired, by multiplying $P_\alpha(F)$ by $\mu^{\alpha-1}$, where $\mu$ is mean income. It is easy to see that this procedure is equivalent to one in which all incomes are normalized by their mean, and (3) is subsequently applied.

Importance of the IA Structure

The theorem represents a particularly sharp characterization of the class of polarization measures that satisfy both the axioms we have imposed and the IA structure. It must be emphasized that both these factors play a role in pinning down our functional form. In fact, it can be checked that several other measures of polarization satisfy Axioms 1–4, though we omit this discussion for the sake of brevity. The IA framework is, therefore, an essential part of the argument.

Partial Ordering

At the same time, and despite the sharpness of the functional form, notice that we do not obtain a complete ordering for polarization, nor do we attempt to do this.\textsuperscript{8} A range of values of $\alpha$ is entertained in the theorem. The union of the complete orderings generated by each value gives us a partial order for polarization. Pinning down this order completely is an open question.

\textsuperscript{8}Indeed, it is possible to impose additional requirements (along the lines explored by ER, for instance) to place narrower bounds on $\alpha$. But we do not consider this necessarily desirable. For instance, the upper value $\alpha = 1$ has the property that all $\lambda$-squeezes of any distribution leave polarization unchanged. We do not feel that a satisfactory measure must possess this feature. This is the reason we are more comfortable with a possible range of acceptable values for $\alpha$. 

A full behavioral foundation for the identification postulate is not within the scope of this paper. However, we make two remarks on the particular specification used here.

First, our axioms imply that identification increases with group size. A well-known problem of collective action (due to Pareto (1906) and Olson (1965)) suggests, however, that smaller groups may sometimes be more effective than larger groups in securing their ends. This argument has been explored by several authors, but perhaps most relevant to the current discussion is Esteban and Ray (2001), which shows that if social conflict arises over the provision of public goods (or even if the good is partially private but the cost function for the supply of lobbying resources has sufficient curvature9), then larger groups are more effective in the aggregate, even though each individual in such groups may be less active owing to the free-rider problem. This finding is consistent with our implication that identification increases with group size.

Second, we remark on our choice of basing identification on the point density. We may more generally suppose that individuals possess a “window of identification” as in ER, Section 4. Individuals within this window would be considered “similar”—possibly with weights decreasing with the distance—and would contribute to a sense of group identity. At the same time, individuals would feel alienated only from those outside the window. Thus, broadening one’s window of identification has two effects. First, it includes more neighbors when computing one’s sense of identification. Second, it reduces one’s sense of distance with respect to aliens—because the width of the identification window affects the “starting point” for alienation.

These two effects can be simultaneously captured in our seemingly narrower model. Let t be some parameter representing the “breadth” in identification. Suppose that this means that each individual x will consider an individual with income y to be at the point \((1 - t)x + ty\). (Thus t is inversely proportional to “breadth”.) The “perceived density” of y from the vantage point of an individual located at x is then

\[
\frac{1}{t} f\left(\frac{y - (1 - t)x}{t}\right)
\]

so that if \(t < 1\), the sense of identification is generally heightened (simply set \(x = y\) above). Thus a small value of t stands for greater identification.

It can be easily shown that the polarization measure resulting from this extended notion of identification is proportional to our measure by the factor \(t^{1-\alpha}\). Therefore, broadening the sense of identification simply amounts to a

9Esteban and Ray prove that if the cost function has enough curvature so that it is “at least” quadratic, then larger groups are more effective even if the conflict is over purely private goods.
re-scaling of the measure defined for the limit case in which one is identified with individuals having exactly the same income.

It is also possible to directly base identification on the average density over a nondegenerate window. It can be shown that when our polarization measure is rewritten to incorporate this notion of identification, it converges precisely to the measure in Theorem 1 as the size of the window converges to zero. Thus an alternative view of point identification is that it is a robust approximation to “narrow” identification windows.

Asymmetric Alienation

In ER we already pointed out that in some environments our implicit hypothesis of a symmetric sense of alienation might not be appropriate. It can be argued that while individuals may feel alienated with respect to those with higher income or wealth, such sentiments need not be reciprocated. For the extreme case of purely one-sided alienation the appropriate extension would be

\[ P_\alpha(f) = \int f(x)^{1+\alpha} \int_x f(y)(y-x) \, dy \, dx. \]

(This is not to say that we have axiomatized such an extension.)

This approach would create a change in the polarization ordering, and depending on the context, it may be a change worth exploring further. The main difference is that (relative to the symmetric case) larger humps or spikes at the lower end of the wealth distribution will be given more weight. In particular, maximal polarization would not be achieved at some symmetric bimodal distribution but at some bimodal distribution that exhibits a larger (local) mode at the bottom of the distribution. This issue is discussed in more detail in ER.

Remarks on the Proof, and the Derived Bounds on \( \alpha \)

The proof of Theorem 1 is long and involved, so a brief roadmap may be useful here. The first half of the proof shows that our axioms imply (3), along with the asserted bounds on \( \alpha \). We begin by noting that the function \( T \) must be (weakly) concave in alienation (Lemmas 1 and 2). Axiom 2 yields this. Yet by Lemmas 3 and 4 (which centrally employ Axiom 3), \( T \) must be (weakly) convex as well. These two assertions must imply that \( T \) is linear in alienation, and so is of the form \( T(i, a) = \phi(i)a \) for some function \( \phi \) (Lemma 4 again). Lemma 5 completes the derivation of our functional form by using the population invariance principle (Axiom 4) to argue that \( \phi \) must exhibit constant elasticity.

Our measure bears an interesting resemblance to the Gini coefficient. Indeed, if \( \alpha = 0 \), the measure is the Gini coefficient. However, our arguments ensure that not only is \( \alpha > 0 \), it cannot go below some uniformly positive lower
bound, which happens to be .25. Where, in the axioms and in the IA structure, does such a bound lurk? To appreciate this, consider Axiom 2, which refers to a double-squeeze of two “side” basic densities. Such squeezes bring down internal alienations in each component density. Yet the axiom demands that overall polarization not fall. It follows, therefore, that the increased identifications created by the squeeze must outweigh the decreased within-component alienation. This restricts $\alpha$. It cannot be too low.

By a similar token, $\alpha$ cannot be too high either. The bite here comes from Axiom 1, which decrees that a single squeeze (in an environment where there is just one basic component) cannot increase polarization. Once again, alienation comes down and some identifications go up (as the single squeeze occurs), but this time we want the decline in alienation to dominate the proceedings. This is tantamount to an upper bound on $\alpha$.\(^{10}\)

The above arguments are made using Lemmas 6 and 7, which also begin the proof that the axioms are implied by our class of measures. The various steps for this direction of the proof, which essentially consist in verifying the axioms, are completed in Lemmas 8–11.

The approach to our characterization bears a superficial similarity to ER. Actually, the axioms are similar in spirit, dealing as they do in each case with issues of identification and alienation. However, their specific structure is fundamentally different. This is because our axioms strongly exploit the density structure of the model (in ER there are only discrete groupings). In turn, this creates basic differences in the method of proof. It is comforting that the two approaches yield the same functional characterization in the end, albeit with different numerical restrictions on the value of $\alpha$.

### 2.5. Comparing Distributions

The fundamental hypothesis underlying all of our analysis is that polarization is driven by the interplay of two forces: identification with one’s own group and alienation vis-a-vis others. Our axioms yield a particular functional form to the interaction between these two forces. When comparing two distributions, which should we expect to display the greater polarization? Our informal answer is that this should depend on the separate contributions of alienation and identification and on their joint co-movement. Increased alienation is associated with an increase in income distances. Increased identification would manifest itself in a sharper definition of groups, i.e., the already highly populated points in the distribution becoming even more populated at the expense of the less populated. Such a change would produce an increase in the variability

\(^{10}\)One might ask: why do the arguments in this paragraph and the one just before lead to “compatible” thresholds for $\alpha$? The reason is this: in the double-squeeze, there are cross-group alienations as well which permit a given increase in identification to have a stronger impact on polarization. Therefore the required threshold on $\alpha$ is smaller in this case.
of the density over the support of the distribution. Finally, when taken jointly, these effects may reinforce each other in the sense that alienation may be highest at the incomes that have experienced an increase in identification, or they may counterbalance each other.

To be sure, it is not possible to move these three factors around independently. After all, one density describes the income distribution and the three factors we have mentioned are byproducts of that density. Nevertheless, thinking in this way develops some intuition for polarization, which we will try to put to use in Section 3.2.

To pursue this line of reasoning, first normalize all incomes by their mean to make the results scale free. Fix a particular value of \( \alpha \), as given by Theorem 1 (more on this parameter below). The \( \alpha \)-identification at income \( y \), denoted by \( \iota_\alpha(y) \), is measured by \( f(y)^\alpha \). Hence, the average \( \alpha \)-identification \( \bar{\iota} \) is defined by

\[
\bar{\iota}_\alpha \equiv \int f(y)^\alpha dF(y) = \int f(y)^{1+\alpha} dy.\tag{4}
\]

The alienation between two individuals with incomes \( y \) and \( x \) is given by \( |y - x| \). Therefore, the overall alienation felt by an individual with income \( y \), \( a(y) \), is

\[
a(y) = \int |y - x| dF(x) \tag{5}
\]

and the average alienation \( \bar{a} \) is

\[
\bar{a} = \int a(y) dF(y) = \int \int |y - x| dF(x) dF(y).\tag{6}
\]

(Notice that \( \bar{a} \) is twice the Gini coefficient.) Now conduct a completely routine exercise. Define \( \rho \) as the normalized covariance between identification and alienation: \( \rho \equiv \text{cov}_{\iota_\alpha, a}/\bar{\iota}_\alpha \bar{a} \). Then

\[
\rho \equiv \frac{\text{cov}_{\iota_\alpha, a}}{\bar{\iota}_\alpha \bar{a}} = \frac{1}{\bar{\iota}_\alpha \bar{a}} \int [\iota_\alpha(y) - \bar{\iota}_\alpha][a(y) - \bar{a}] f(y) dy
\]

\[
= \frac{1}{\bar{\iota}_\alpha \bar{a}} \left[ \int f(y)^{1+\alpha} a(y) dy - \bar{a} \bar{\iota}_\alpha \right]
\]

\[
= \frac{P_\alpha(f)}{\bar{\iota}_\alpha \bar{a}} - 1,
\]

so that

\[
P_\alpha(f) = \bar{a} \bar{\iota}_\alpha [1 + \rho].\tag{7}
\]
This is a more precise statement of the informal idea expressed at the start of this section. There is one dimension, however, along which this decomposition lacks intuition. It is that $\alpha$ unavoidably enters into it: we make this explicit by using the term $\alpha$-identification (though we will resort to “identification” when there is little risk of confusion). This sort of identification is not intrinsic to the density. Yet the formula itself is useful, for it tells us that—all other things being equal—greater variation, “spikiness,” or multimodality in the density is likely to translate into greater polarization for that density, this effect making itself felt more strongly when $\alpha$ is larger. The reason is simple: the main ingredient for $\alpha$-identification is the function $x^{1+\alpha}$ (see (4)), which is a strictly convex function of $x$.

The connection with spikiness or multimodality ties in with our graphical intuitions regarding polarization. We reiterate, however, that this is only one factor of several, and that often it may not be possible to change this factor in the direction of higher polarization without infringing the ceteris paribus qualification. For instance, if a unimodal density is altered by the introduction of two or more local modes, such multimodality per se may not bring higher polarization with it. This is because the existence of several modes may also bring average alienation down. In particular, a highly skewed distribution with a single mode may still exhibit greater polarization relative to other bimodal distributions. Nevertheless, the connection with “variability” may be helpful in some situations, and we will invoke it in the empirical discussion of Section 3.2. Indeed, in unimodal situations (which present the most subtle problems as far as polarization is concerned), these factors can act as guides to simple visual inspection.

3. ESTIMATION AND ILLUSTRATION

3.1. Estimation and Statistical Inference

We now turn to estimation issues regarding $P_\alpha(F)$ and associated questions of statistical inference. The details of this discussion have been omitted to economize on space, but may be found in the working paper version of this paper (Duclos, Esteban, and Ray (2003)). First note that for every distribution function $F$ with associated density $f$ and mean $\mu$, we have that

\begin{equation}
P_\alpha(F) = \int_y f(y)^\alpha a(y) dF(y),
\end{equation}

with $a(y) \equiv \mu + y(2F(y) - 1) - 2 \int_{-\infty}^y x dF(x)$. Suppose, then, that we wish to estimate $P_\alpha(F)$ using a random sample of $n$ i.i.d. observations of in-
come \( y_i, i = 1, \ldots, n \), drawn from the distribution \( F(y) \) and ordered such that \( y_1 \leq y_2 \leq \cdots \leq y_n \). A natural estimator of \( P_\alpha(F) \) is

\[
P_\alpha(\hat{F}) = n^{-1} \sum_{i=1}^{n} \hat{f}(y_i)^\alpha \hat{a}(y_i),
\]

where \( \hat{a}(y_i) \) is given as

\[
\hat{a}(y_i) = \hat{\mu} + y_i \left( n^{-1} (2i - 1) - 1 - \frac{1}{2} \sum_{j=1}^{i-1} y_j + y_i \right),
\]

\( \hat{\mu} \) is the sample mean, and where \( \hat{f}(y_i)^\alpha \) is estimated nonparametrically using kernel estimation procedures.\(^{11} \) These procedures use a symmetric kernel function \( K(u) \), defined such that \( \int_{-\infty}^{\infty} K(u) \, du = 1 \) and \( K(u) \geq 0 \)—a Gaussian kernel is used in the illustration. The estimator \( \hat{f}(y) \) is then defined as \( \hat{f}(y) = n^{-1} \sum_{i=1}^{n} K_h(y - y_i) \), with \( K_h(z) \equiv h^{-1} K(z/h) \) and \( h \) being a bandwidth parameter. A common technique to select an “optimal” bandwidth \( h^* \) is to minimize the mean square error (MSE) of the estimator, given a sample of size \( n \). A “rule-of-thumb” formula that can be used to do this in our context is approximately given by

\[
h^* \approx 4.7 n^{-5/2} \sigma_{\ln}^{-1}.
\]

Easily computed, this formula works well with the normal distribution since it is then never farther than 5% from the \( h^* \) that truly minimizes the MSE. For skewness larger than about 6, a more robust—though more cumbersome—approximate formula for the computation of \( h^* \) is given by

\[
h^* \approx n^{-5/2} IQ \frac{(3.76 + 14.7 \sigma_{\ln})}{(1 + 1.09 \cdot 10^{-4} \sigma_{\ln})^{(7268+15323\alpha)}}
\]

where \( IQ \) is the interquartile and \( \sigma_{\ln} \) is the variance of the logarithms of income.

It can also be shown (under certain mild regularity conditions) that \( n^5 (P_\alpha(\hat{F}) - P_\alpha(F)) \) has an asymptotic limiting normal distribution \( N(0, V_\alpha) \), with

\[
V_\alpha = \text{var}_{f(y)} \left( (1 + \alpha) f(y)^\alpha a(y) + y \int f(x)^\alpha \, dF(x) 
+ 2 \int_y^\infty (x - y) f(x)^\alpha \, dF(x) \right).
\]

\(^{11} \)The literature on kernel density estimation is large—see for instance Silverman (1986), Härdle (1990), and Pagan and Ullah (1999) for an introduction to it.
This result is distribution-free in the sense that everything in the above can be estimated consistently without having to specify the population distribution from which the sample is drawn.

3.2. An Illustration

We illustrate the above results with data drawn from the Luxembourg Income Study (LIS) data sets on 21 countries for each of Wave 3 (1989–1992) and Wave 4 (1994–1997). Countries, survey years, and abbreviations are listed in Table I. We use household disposable income (i.e., post-tax-and-transfer income) normalized by an adult-equivalence scale defined as $s^5$, where $s$ is household size. Observations with negative incomes are removed as well as those with incomes exceeding 50 times the average (this affects less than 1% of all samples). Household observations are weighted by the LIS sample weights times the number of persons in the household. As discussed in Section 2.4, the usual homogeneity-of-degree-zero property is imposed throughout by multiplying the indices $P_\alpha(F)$ by $\mu^{\alpha-1}$ or equivalently by normalizing all incomes by

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Countries</th>
<th>Years</th>
<th>Sample Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>as</td>
<td>Australia</td>
<td>1989/1994</td>
<td>16,331/7,441</td>
</tr>
<tr>
<td>be</td>
<td>Belgium</td>
<td>1992/1997</td>
<td>3,821/4,632</td>
</tr>
<tr>
<td>cn</td>
<td>Canada</td>
<td>1991/1994</td>
<td>21,647/40,849</td>
</tr>
<tr>
<td>dk</td>
<td>Denmark</td>
<td>1992/1995</td>
<td>12,895/13,124</td>
</tr>
<tr>
<td>fi</td>
<td>Finland</td>
<td>1991/1995</td>
<td>11,749/9,263</td>
</tr>
<tr>
<td>fr</td>
<td>France</td>
<td>1989/1994</td>
<td>9,038/11,294</td>
</tr>
<tr>
<td>ge</td>
<td>Germany</td>
<td>1989/1994</td>
<td>4,187/6,045</td>
</tr>
<tr>
<td>hu</td>
<td>Hungary</td>
<td>1991/1994</td>
<td>2,019/1,992</td>
</tr>
<tr>
<td>is</td>
<td>Israel</td>
<td>1992/1997</td>
<td>5,212/5,230</td>
</tr>
<tr>
<td>it</td>
<td>Italy</td>
<td>1991/1994</td>
<td>8,188/8,135</td>
</tr>
<tr>
<td>lx</td>
<td>Luxembourg</td>
<td>1991/1994</td>
<td>1,957/1,813</td>
</tr>
<tr>
<td>mx</td>
<td>Mexico</td>
<td>1989/1996</td>
<td>11,531/14,042</td>
</tr>
<tr>
<td>nw</td>
<td>Norway</td>
<td>1991/1995</td>
<td>8,073/10,127</td>
</tr>
<tr>
<td>pl</td>
<td>Poland</td>
<td>1992/1995</td>
<td>6,602/32,009</td>
</tr>
<tr>
<td>rc</td>
<td>Rep. of China / Taiwan</td>
<td>1991/1995</td>
<td>16,434/14,706</td>
</tr>
<tr>
<td>ru</td>
<td>Russia</td>
<td>1992/1995</td>
<td>6,361/3,518</td>
</tr>
<tr>
<td>sw</td>
<td>Sweden</td>
<td>1992/1995</td>
<td>12,484/16,260</td>
</tr>
<tr>
<td>uk</td>
<td>United Kingdom</td>
<td>1991/1995</td>
<td>7,056/6,797</td>
</tr>
<tr>
<td>us</td>
<td>United States</td>
<td>1991/1994</td>
<td>16,052/66,014</td>
</tr>
</tbody>
</table>

12See http://lissy.ceps.lu for detailed information on the structure of these data.
their mean. For ease of comparison, all indices are divided by 2, so that \( P_{\alpha=0}(F) \) is the usual Gini coefficient.

Tables II and III show estimates of the Gini (\( P_0 \)) and four polarization indices (\( P_\alpha \) for \( \alpha = .25, .5, .75, 1 \)) in 21 countries for each of the two waves, along with their asymptotic standard deviations. The polarization indices are typically rather precisely estimated, with often only the third decimal of the estimators being subject to sampling variability. Using a conventional test size of 5%, it can be checked that around 90% of the possible cross-country comparisons are statistically significant, whatever the value of \( \alpha \). Tables II and III also show the country rankings, with a high rank corresponding to a relatively large value of the relevant index, and with countries displayed by their order in the Gini ranking.

**Polarization Behaves Differently from Inequality**

Observe first that \( P_0 \) and \( P_{0.25} \) induce very similar rankings. But considerable differences arise between \( P_0 \) and \( P_1 \). For instance, for Wave 3, the Czech Republic has the lowest Gini index of all countries, but ranks 11 in terms of \( P_1 \). Conversely, Canada, Australia, and the United States exhibit high Gini inequality, but relatively low \("P_1\)-polarization." The correlation across country rankings for different \( \alpha \)'s clearly falls as the distance between the \( \alpha \)'s increases. The lowest Pearson correlation of all—.6753—is the correlation between the Gini index and \( P_1 \) in Wave 3. Clearly, polarization and inequality are naturally correlated, but they are also empirically distinct in this dataset. Moreover, the extent to which inequality comparisons resemble polarization comparisons depends on the parameter \( \alpha \), which essentially captures the power of the identification effect.

**Alienation and Identification**

Recall the decomposition exercise carried out in Section 2.5, in which we obtained (7), reproduced here for convenience:

\[
P_\alpha = \bar{a}\bar{\iota}_\alpha[1 + \rho].
\]

Table IV summarizes the relevant statistics for all Wave 3 countries, decomposing polarization as the product of average alienation, average identification, and (one plus) the normalized covariance between the two. Consider \( \alpha = 1 \). Note that the bulk of cross-country variation in polarization stems from significant variation in average identification as well as in average alienation. In contrast, the covariance between the two does not exhibit similar variation across countries. Some countries (Finland, Sweden, and Denmark) rank low both in terms of inequality and polarization, due in large part to low average alienation. Some countries, most strikingly Russia, Mexico, and the UK rank consistently high both in terms of inequality and polarization—even though average
<table>
<thead>
<tr>
<th>Country</th>
<th>(\alpha = 0)</th>
<th>(\alpha = .25)</th>
<th>(\alpha = .50)</th>
<th>(\alpha = .75)</th>
<th>(\alpha = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index Rkg</td>
<td>Index Rkg</td>
<td>Index Rkg</td>
<td>Index Rkg</td>
<td>Index Rkg</td>
</tr>
<tr>
<td>cz92</td>
<td>.2082 1 .0023</td>
<td>.1767 1 .0014</td>
<td>.1637 2 .0011</td>
<td>.1585 4 .0011</td>
<td>.1575 11 .0012</td>
</tr>
<tr>
<td>fi91</td>
<td>.2086 2 .0017</td>
<td>.1782 2 .0010</td>
<td>.1611 1 .0007</td>
<td>.1505 1 .0005</td>
<td>.1436 1 .0005</td>
</tr>
<tr>
<td>sw92</td>
<td>.2267 4 .0019</td>
<td>.1888 3 .0012</td>
<td>.1674 3 .0008</td>
<td>.1543 2 .0006</td>
<td>.1459 2 .0006</td>
</tr>
<tr>
<td>nw91</td>
<td>.2315 5 .0029</td>
<td>.1919 5 .0017</td>
<td>.1713 5 .0013</td>
<td>.1588 5 .0011</td>
<td>.1505 5 .0011</td>
</tr>
<tr>
<td>ge89</td>
<td>.2469 8 .0048</td>
<td>.2019 8 .0028</td>
<td>.1779 7 .0021</td>
<td>.1634 7 .0020</td>
<td>.1540 7 .0021</td>
</tr>
<tr>
<td>is92</td>
<td>.3055 16 .0018</td>
<td>.2421 17 .0011</td>
<td>.2051 17 .0008</td>
<td>.1804 18 .0006</td>
<td>.1626 18 .0006</td>
</tr>
<tr>
<td>ru92</td>
<td>.4017 20 .0019</td>
<td>.2957 20 .0012</td>
<td>.2400 20 .0008</td>
<td>.2046 20 .0006</td>
<td>.1797 20 .0006</td>
</tr>
<tr>
<td></td>
<td>.0055</td>
<td>.0034</td>
<td>.0030</td>
<td>.0032</td>
<td>.0036</td>
</tr>
</tbody>
</table>

Note: Standard errors appear on every second line.


<table>
<thead>
<tr>
<th>Country</th>
<th>$\alpha = 0$</th>
<th>Index</th>
<th>Rkg</th>
<th>$\alpha = .25$</th>
<th>Index</th>
<th>Rkg</th>
<th>$\alpha = .50$</th>
<th>Index</th>
<th>Rkg</th>
<th>$\alpha = .75$</th>
<th>Index</th>
<th>Rkg</th>
<th>$\alpha = 1$</th>
<th>Index</th>
<th>Rkg</th>
</tr>
</thead>
<tbody>
<tr>
<td>fi95</td>
<td>.2174</td>
<td>1</td>
<td>.1832</td>
<td>1</td>
<td>.1661</td>
<td>2</td>
<td>.1564</td>
<td>2</td>
<td>.1506</td>
<td>6</td>
<td>.0027</td>
<td>.0016</td>
<td>.0012</td>
<td>.0011</td>
<td>.0012</td>
</tr>
<tr>
<td>sw95</td>
<td>.2218</td>
<td>2</td>
<td>.1845</td>
<td>2</td>
<td>.1652</td>
<td>1</td>
<td>.1549</td>
<td>1</td>
<td>.1498</td>
<td>3</td>
<td>.0019</td>
<td>.0012</td>
<td>.0008</td>
<td>.0007</td>
<td>.0008</td>
</tr>
<tr>
<td>hr94</td>
<td>.2353</td>
<td>3</td>
<td>.1978</td>
<td>4</td>
<td>.1764</td>
<td>4</td>
<td>.1633</td>
<td>7</td>
<td>.1549</td>
<td>8</td>
<td>.0043</td>
<td>.0028</td>
<td>.0021</td>
<td>.0017</td>
<td>.0019</td>
</tr>
<tr>
<td>be97</td>
<td>.2496</td>
<td>5</td>
<td>.2061</td>
<td>5</td>
<td>.1796</td>
<td>5</td>
<td>.1616</td>
<td>4</td>
<td>.1486</td>
<td>1</td>
<td>.0029</td>
<td>.0018</td>
<td>.0012</td>
<td>.0010</td>
<td>.0010</td>
</tr>
<tr>
<td>nl94</td>
<td>.2558</td>
<td>7</td>
<td>.2094</td>
<td>7</td>
<td>.1812</td>
<td>7</td>
<td>.1624</td>
<td>5</td>
<td>.1491</td>
<td>2</td>
<td>.0029</td>
<td>.0018</td>
<td>.0012</td>
<td>.0009</td>
<td>.0010</td>
</tr>
<tr>
<td>hu94</td>
<td>.3248</td>
<td>15</td>
<td>.2486</td>
<td>15</td>
<td>.2087</td>
<td>15</td>
<td>.1852</td>
<td>15</td>
<td>.1700</td>
<td>18</td>
<td>.0081</td>
<td>.0048</td>
<td>.0037</td>
<td>.0035</td>
<td>.0038</td>
</tr>
<tr>
<td>is97</td>
<td>.3371</td>
<td>16</td>
<td>.2598</td>
<td>17</td>
<td>.2159</td>
<td>17</td>
<td>.1871</td>
<td>18</td>
<td>.1666</td>
<td>17</td>
<td>.0044</td>
<td>.0025</td>
<td>.0019</td>
<td>.0018</td>
<td>.0020</td>
</tr>
<tr>
<td>us94</td>
<td>.3622</td>
<td>19</td>
<td>.2747</td>
<td>19</td>
<td>.2223</td>
<td>19</td>
<td>.1868</td>
<td>17</td>
<td>.1610</td>
<td>11</td>
<td>.0010</td>
<td>.0006</td>
<td>.0004</td>
<td>.0004</td>
<td>.0004</td>
</tr>
<tr>
<td>ru95</td>
<td>.4497</td>
<td>20</td>
<td>.3222</td>
<td>20</td>
<td>.2566</td>
<td>20</td>
<td>.2164</td>
<td>20</td>
<td>.1889</td>
<td>20</td>
<td>.0061</td>
<td>.0035</td>
<td>.0028</td>
<td>.0028</td>
<td>.0030</td>
</tr>
</tbody>
</table>

Note: Standard errors appear on every second line.
### TABLE IV
ALIENATION AND IDENTIFICATION—LIS WAVE 3

<table>
<thead>
<tr>
<th>Country</th>
<th>$\alpha = 0$</th>
<th>$\alpha = .25$</th>
<th>$\alpha = .50$</th>
<th>$\alpha = .75$</th>
<th>$\alpha = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\bar{i}$</td>
<td>$c$</td>
<td>$\bar{i} \cdot c$</td>
<td>$P$</td>
<td>$\bar{i}$</td>
</tr>
<tr>
<td>as89</td>
<td>.3084</td>
<td>.8508</td>
<td>.9227</td>
<td>.7851</td>
<td>.2421</td>
</tr>
<tr>
<td>be92</td>
<td>.2233</td>
<td>.9110</td>
<td>.9327</td>
<td>.8497</td>
<td>.1897</td>
</tr>
<tr>
<td>cz92</td>
<td>.2891</td>
<td>.8634</td>
<td>.9219</td>
<td>.7960</td>
<td>.2301</td>
</tr>
<tr>
<td>dk92</td>
<td>.2081</td>
<td>.9504</td>
<td>.8935</td>
<td>.8492</td>
<td>.1767</td>
</tr>
<tr>
<td>fr91</td>
<td>.2367</td>
<td>.9051</td>
<td>.9169</td>
<td>.8298</td>
<td>.1964</td>
</tr>
<tr>
<td>ge89</td>
<td>.2815</td>
<td>.8782</td>
<td>.9015</td>
<td>.7917</td>
<td>.2229</td>
</tr>
<tr>
<td>hu91</td>
<td>.2469</td>
<td>.9021</td>
<td>.9066</td>
<td>.8179</td>
<td>.2019</td>
</tr>
<tr>
<td>is92</td>
<td>.3055</td>
<td>.8626</td>
<td>.9188</td>
<td>.7926</td>
<td>.2421</td>
</tr>
<tr>
<td>it91</td>
<td>.2887</td>
<td>.8676</td>
<td>.9212</td>
<td>.7993</td>
<td>.2307</td>
</tr>
<tr>
<td>mx89</td>
<td>.4909</td>
<td>.8343</td>
<td>.8453</td>
<td>.7052</td>
<td>.3462</td>
</tr>
<tr>
<td>nl91</td>
<td>.2633</td>
<td>.8952</td>
<td>.9003</td>
<td>.8059</td>
<td>.2122</td>
</tr>
<tr>
<td>nw91</td>
<td>.2315</td>
<td>.9128</td>
<td>.9082</td>
<td>.8290</td>
<td>.1919</td>
</tr>
<tr>
<td>pl27</td>
<td>.2737</td>
<td>.8837</td>
<td>.9067</td>
<td>.8013</td>
<td>.2193</td>
</tr>
<tr>
<td>re91</td>
<td>.2708</td>
<td>.8883</td>
<td>.9099</td>
<td>.8083</td>
<td>.2188</td>
</tr>
<tr>
<td>ru92</td>
<td>.4017</td>
<td>.8300</td>
<td>.8868</td>
<td>.7361</td>
<td>.2957</td>
</tr>
<tr>
<td>sw92</td>
<td>.2267</td>
<td>.9077</td>
<td>.9177</td>
<td>.8330</td>
<td>.1888</td>
</tr>
<tr>
<td>uk91</td>
<td>.3381</td>
<td>.8521</td>
<td>.9047</td>
<td>.7709</td>
<td>.2607</td>
</tr>
<tr>
<td>us91</td>
<td>.3394</td>
<td>.8298</td>
<td>.9320</td>
<td>.7734</td>
<td>.2625</td>
</tr>
</tbody>
</table>
identification for the three countries is among the lowest of all. Average alienation is very high in these countries. Yet other countries show low inequality but relatively high polarization, while others exhibit the reverse relative rankings.

While maintaining the same average alienation, the UK density exhibits higher variability than its US counterpart—see the upper panel in Figure 4. The US distribution shows a remarkably flat density on the interval \([0.25, 1.25]\) of normalized incomes and so has thick tails. In contrast, the UK displays a clear mode at \(y = 0.4\) and thinner tails. Because—as already discussed in Section 2.5—the identification function \(f'\) times the density \(f\) is strictly convex in \(f\), the country with the greater variation in identification will exhibit a higher value of average identification, with the difference growing more pronounced as \(\alpha\) increases. To be sure, variations in identification find their starkest expression when distributions are multimodal, but even without such multimodality, variation is possible.

Remember that variation in identification is only one of several factors. In particular, we do not mean to suggest that the country with the greater variation in identification will invariably exhibit greater polarization as \(\alpha \to 1\). For instance, our notion of a squeeze increases the variability of identification, but polarization must fall, by Axiom 1 (this is because alienation falls too with the squeeze). See the discussion at the end of Section 2.5 for another illustration of this point.

**Sensitivity to \(\alpha\)**

As \(\alpha\) increases from .25 to 1, the cross-country variation in the value of average \(\alpha\)-identification goes up. This increase in cross-country variability produces frequent “crossings” in the ranking of countries by polarization. Such crossings can occur at very low values of \(\alpha\) (below .25) so that for all \(\alpha \in [0.25, 1]\) the polarization ranking opposes the inequality ranking. This is the case (for Wave 3) for Belgium–Sweden, Italy–Canada, and Israel–Australia. Crossings could—and do—occur for intermediate values of \(\alpha \in [0.25, 1]\). To be sure, they may not occur for any \(\alpha \leq 1\), thus causing the polarization ordering to coincide with the inequality ordering. This is indeed a most frequent case for pairwise comparisons in Wave 3. Finally, in Wave 4 we also observe “double crossings” in the cases of Canada–France and Australia–Poland. In both cases the first country starts with higher inequality, \(P_0\), followed by a lower value of \(P_{25}\), but later returning to higher values for larger values of \(\alpha\).

UK inequality is very close to US inequality; for all intents and purposes the two have the same Gini in Wave 3. Indeed, the UK ranks eighteenth and the US nineteenth—this is true for any \(\alpha < 0.33\). However, as \(\alpha\) increases beyond .33 up to 1, the UK retains the nineteenth position, while the US descends to ninth in the rankings. That fall in rankings occurs mostly when \(\alpha\) increases beyond .75.

The Czech–US densities provide additional visual support to this sensitivity—see the lower panel in Figure 4. Here, the basic inequality comparison
FIGURE 4.—Estimated densities for the US, UK, and Czech Republic, Wave 3.
is unambiguous: the Czech Republic has lower inequality than the US. But the Czech Republic has a spikier density with greater variation in it. This “shadow of multimodality” kicks in as $\alpha$ is increased, so much so that the Czech Republic is actually deemed equally or more polarized than the US by the time $\alpha = 1$.

Partial Ordering

One might respond to the above observations as follows: our axiomatics do not rule out values of $\alpha$ very close to .25. Hence, in the strict sense of a partial order we are unable to (empirically) distinguish adequately between inequality and polarization, at least with the dataset at hand. In our opinion this response would be too hasty. Our characterization not only implies a partial ordering, it provides a very clean picture of how that ordering is parameterized, with the parameter $\alpha$ having a definite interpretation. If substantial variations in ranking occur as $\alpha$ increases, this warrants a closer look, and certainly shows—empirically—how “large” subsets of polarization indices work very differently from the Gini inequality index.

4. FINAL REMARKS, AND A PROPOSED EXTENSION

In this paper we present and characterize a class of measures for income polarization, based on what we call the identification-alienation structure. Our approach is fundamentally based on the view that interpersonal alienation fuels a polarized society, as does inequality. Our departure from inequality measurement lies in the notion that such alienation must also be complemented by a sense of identification. This combination of the two forces generates a class of measures that are sensitive (in the same direction) to both elements of inequality and equality, depending on where these changes are located in the overall distribution.

Our characterization, and the alternative decomposition presented in (7), permit us to describe the measure very simply: for any income distribution, polarization is the product of average alienation, average identification, and (one plus) the mean-normalized covariance between these two variables. We also discuss estimation issues for our measures as well as associated questions of statistical inference.

We wish to close this paper with some remarks on what we see to be the main conceptual task ahead. Our analysis generates a certain structure for identification and alienation functions in the special case in which both identification and alienation are based on the same characteristic. This characteristic can be

13One would expect these distinctions to magnify even further for distributions that are not unimodal (unfortunately, this exploration is not permitted by our dataset). For instance, one might use our measures to explore the “twin-peaks” property identified by Quah (1996) for the world distribution of income. But this is the subject of future research.
income or wealth. In principle it could be any measurable feature with a well-defined ordering. The key restriction, however, is that whatever we choose the salient characteristic for identification to be, intergroup alienation has to be driven by the very same characteristic. This seems obvious in the cases of income or wealth. Yet, for some relevant social characteristics this might not be a natural assumption. Think of the case of ethnic polarization. It may or may not seem appropriate here to base interethnic alienation as only depending on some suitably defined “ethnicity distance.” In the cases of socially based group identification we find it more compelling to adopt a multidimensional approach to polarization, permitting alienation to depend on characteristics other than the one that defines group identity. In this proposed extension, we liberally transplant our findings to the case of social polarization, but with no further axiomatic reasoning. In our opinion, such reasoning is an important subject of future research.

Suppose, then, that there are $M$ “social groups,” based on region, kin, ethnicity, religion, . . . Let $n_j$ be the number of individuals in group $j$, with overall population normalized to one. Let $F_j$ describe the distribution of income in group $j$ (with $f_j$ the accompanying density), unnormalized by group population. One may now entertain a variety of “social polarization measures.”

### 4.1. Pure Social Polarization

Consider, first, the case of “pure social polarization,” in which income plays no role. Assume that each person is “fully” identified with every other member of his group. Likewise, the alienation function takes on values that are specific to group pairs and have no reference to income. For each pair of groups $j$ and $k$ denote this value by $\Delta_{jk}$. Then a natural transplant of (3) yields the measure

$$P_s(F) = \sum_{j=1}^{M} \sum_{k=1}^{M} \alpha_{jk} n_j n_k \Delta_{jk}. \quad (14)$$

Even this sort of specification may be too general in some interesting instances in which individuals are interested only in the dichotomous perception Us/They. In particular, in these instances, individuals are not interested in differentiating between the different opposing groups. Perhaps the simplest instance of this is a pure contest (Esteban and Ray (1999)), which yields the variant\textsuperscript{14}

$$\tilde{P}_s(F) = \sum_{j=1}^{M} n_j^{1+\alpha} (1 - n_j). \quad (15)$$

\textsuperscript{14}See Reynal-Querol (2002) for a similar analysis. D’Ambrosio and Wolff (2001) also consider a measure of this type but with income distances across groups explicitly considered.
4.2. Hybrids

Once the two extremes—pure income polarization and pure social polarization—are identified, we may easily consider several hybrids. As examples, consider the case in which notions of identification are mediated not just by group membership but by income similarities as well, while the antagonism equation remains untouched. Then we get what one might call social polarization with income-mediated identification:

\[ P_s(F) = \sum_{j=1}^{M} (1 - n_j) \int_x f_j(x)^a \, dF_j(x). \]

One could expand (or contract) the importance of income further, while still staying away from the extremes. For instance, suppose that—in addition to the income-mediation of group identity—alienation is also income-mediated (for alienation, two individuals must belong to different groups and have different incomes). Now groups have only a demarcating role—they are necessary (but not sufficient) for identity, and they are necessary (but not sufficient) for alienation. The resulting measure would look like this:

\[ P^*(F) = \sum_{j=1}^{M} \sum_{k \neq j} \int_x \int_y f_j(x)^a |x - y| \, dF_j(x) \, dF_k(y). \]

Note that we do not intend to suggest that other special cases or hybrids are not possible, or that they are less important. The discussion here is only to show that social and economic considerations can be profitably combined in the measurement of polarization. Indeed, it is conceivable that such measures will perform better than the more commonly used fragmentation measures in the analysis of social conflict. But a full exploration of this last theme must await a future paper.

Département d’économique and CIRPÉÉ, Pavillon J.-A.-DeSève, Université Laval, Québec, Canada G1K 7P4; jyves@ecn.ulaval.ca,

Instituto de Análisis Económico, Universidad Autónoma de Barcelona, 08193 Bellaterra, Barcelona, Spain, and Universitat Pompeu Fabra, Barcelona, Spain; Joan.Esteban@uab.es,

and

Dept. of Economics, New York University, 269 Mercer Street, New York, NY 10003, U.S.A.; debraj.ray@nyu.edu; http://www.econ.nyu.edu/user/debraj/.

Manuscript received December, 2002; final revision received May, 2004.

APPENDIX

PROOF OF THEOREM 1: First, we show that Axioms 1–4 imply (3). The lemma below follows from Jensen’s inequality; proof is omitted.
LEMMA 1: Let $g$ be a continuous real-valued function defined on $\mathbb{R}$ such that for all $x > 0$ and all $\delta$ with $0 < \delta < x$,

$$g(x) \geq \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} g(y) \, dy.$$  

Then $g$ must be a concave function.

In what follows, remember that our measure only considers income differences across people, so that we may slide any distribution to left or right as we please.

LEMMA 2: The function $T$ must be concave in $a$ for every $i > 0$.

PROOF: Fix $x > 0$, some $i > 0$, and $\delta \in (0, x)$. Consider three basic densities as in Axiom 2 (see Figure 2) but specialize as shown in Figure 5; each is a transform of a uniform basic density. The bases are centered at $-x$, 0, and $x$. The side densities are of width $2\delta$ and height $h$, and the middle density is of width $2\epsilon$ and height $i$. We shall vary $\epsilon$ and $h$ but to make sure that Axiom 2 applies, we choose $\epsilon > 0$ such that $\delta + \epsilon < x$. A $\lambda$-squeeze of the side densities simply contracts their base width to $2\lambda \delta$, while the height is raised to $h/\lambda$. For each $\lambda$, decompose the measure (1) into five components: (a) the “internal polarization” $P_m$ of the middle rectangle (this component does not vary with $\lambda$ so there will be no need to explicitly calculate it); (b) the “internal polarization” $P_s$ of each side rectangle; (c) total effective antagonism, $A_{ms}$, felt by inhabitants of the middle towards each side density; (d) total effective antagonism, $A_{sm}$, felt by inhabitants of each side towards the middle; (e) total effective antagonism, $A_{ss}$, felt by inhabitants of one side towards the other side. Each of these last four terms appear twice, so that (writing everything as a function of $\lambda$),

$$P(\lambda) = P_m + 2P_s(\lambda) + 2A_{ms}(\lambda) + 2A_{sm}(\lambda) + 2A_{ss}(\lambda).$$

Now we compute the terms on the right-hand side of (19). First,

$$P_s(\lambda) = \frac{1}{\lambda^2} \int_{-\lambda - \delta}^{0} \int_{0}^{\lambda + \delta} T(h/\lambda, |b' - b|) h^2 \, db' \, db,$$

FIGURE 5.
where (here and in all subsequent cases) $b$ will stand for the “origin” income (to which the identification is applied) and $b'$ for the “destination” income (towards which the antagonism is felt). Next,

$$A_{ms}(\lambda) = \frac{1}{\lambda} \int_{-\epsilon}^{\epsilon} \int_{x-\lambda \delta}^{x+\lambda \delta} T(i, b' - b) db' \, db.$$ 

Third,

$$A_{sm}(\lambda) = \frac{1}{\lambda} \int_{x-\lambda \delta}^{x+\lambda \delta} \int_{-\epsilon}^{\epsilon} T(h/\lambda, b - b') db' \, db,$$

and finally,

$$A_{ss}(\lambda) = \frac{1}{\lambda^2} \int_{-\epsilon}^{\epsilon} \int_{x-\lambda \delta}^{x+\lambda \delta} T(h/\lambda, b' - b) db' \, db.$$ 

The axiom requires that $P(\lambda) \geq P(1)$. Equivalently, we require that \left[ P(\lambda) - P(1) \right] / 2h \geq 0$ for all $h$, which implies in particular that

$$\liminf_{h \to 0} \frac{P(\lambda) - P(1)}{2h} \geq 0. \quad (20)$$ 

If we divide through by $h$ in the individual components calculated above and then send $h$ to 0, it is easy to see that the only term that remains is $A_{ms}$. Formally, (20) and the calculations above must jointly imply that

$$\frac{1}{\lambda} \int_{-\epsilon}^{\epsilon} \int_{x-\lambda \delta}^{x+\lambda \delta} T(i, b' - b) \, db' \, db \geq \int_{-\epsilon}^{\epsilon} \int_{x-\delta}^{x+\delta} T(i, b' - b) \, db' \, db, \quad (21)$$

and this must be true for all $\lambda \in (0, 1)$ as well as all $\epsilon \in (0, x - \delta)$. Therefore we may insist on the inequality in (21) holding as $\lambda \to 0$. Performing the necessary calculations, we may conclude that

$$\frac{1}{\epsilon} \int_{-\epsilon}^{\epsilon} T(i, x - b) \, db \geq \frac{1}{\epsilon} \int_{-\epsilon}^{\epsilon} \int_{x-\delta}^{x+\delta} T(i, b' - b) \, db' \, db \quad \text{for all } \epsilon \in (0, x - \delta). \quad (22)$$

Finally, take $\epsilon$ to zero in (22). This allows us to deduce that

$$T(i, x) \geq \int_{x-\delta}^{x+\delta} T(i, b') \, db'. \quad (23)$$

As (23) must hold for every $x > 0$ and every $\delta \in (0, x)$, we may invoke Lemma 1 to conclude that $T$ is concave in $x$ for every $i > 0$. \textit{Q.E.D.}

**Lemma 3:** Let $g$ be a concave, continuous function on $\mathbb{R}_+$, with $g(0) = 0$. Suppose that, for each $a$ and $a'$ with $a > a' > 0$, there exists $\Delta > 0$ such that

$$g(a + \Delta) - g(a) \geq g(a') - g(a' - \Delta) \quad \text{for all } \Delta \in (0, \bar{\Delta}). \quad (24)$$

Then $g$ must be linear.

The proof is straightforward and is omitted.

**Lemma 4:** There is a continuous function $\phi(i)$ such that $T(i, a) = \phi(i)a$ for all $i$ and $a$. 

FIGURE 6.

Proof: Fix \(a\) and \(a'\) with \(a > a' > 0\), and \(i > 0\). Consider four basic densities as in Axiom 3 (see Figure 3) but specialize as shown in Figure 6; each is a transform of a uniform basic density. The bases are centered at locations \(-y, -x, x,\) and \(y\), where \(x \equiv (a - a')/2\) and \(y \equiv (a + a')/2\). The “inner” densities are of width \(2\delta\) and height \(h\), and the “outer” densities are of width \(2\epsilon\) and height \(i\). We shall vary different parameters (particularly \(x\)) but to ensure disjoint support we assume throughout that \(\epsilon < x\) and \(\delta + \epsilon < y - x - \Delta\) for some \(\Delta > 0\). Again, decompose the polarization measure (1) into several distinct components: (a) the “internal polarization” of each rectangle \(j\) (call it \(P_j\), \(j = 1, 2, 3, 4\)); these components are unchanged as we change \(x\) so there will be no need to calculate them explicitly; (b) total effective antagonism \(A_{jk}(x)\) felt by inhabitants of rectangle \(j\) towards rectangle \(k\) (we emphasize dependence on the parameter \(x\)). Thus total polarization \(P(x)\) is given by

\[
P(x) = \sum_{j=1}^{4} P_j + \sum_{j} \sum_{kj} A_{jk}(x)
\]

\[
= \sum_{j=1}^{4} P_j + 2A_{12}(x) + 2A_{13}(x) + 2A_{21}(x) + 2A_{31}(x) + 2A_{23}(x) + 2A_{14},
\]

where the second equality simply exploits obvious symmetries and \(A_{14}\) is noted to be independent of \(x\). Let us compute the terms in this formula that do change with \(x\). We have

\[
A_{12}(x) = \int_{-y-\epsilon}^{-y+\epsilon} \int_{-\delta-x}^{-\delta-x+\delta} T(i, b' - b) i h \ db' \ db,
\]

\[
A_{13}(x) = \int_{-y-\epsilon}^{-y+\epsilon} \int_{x-\delta}^{-x+\delta} T(i, b' - b) i h \ db' \ db,
\]

\[
A_{21}(x) = \int_{-x-\delta}^{-x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b - b') i h \ db' \ db,
\]

\[
A_{23}(x) = \int_{-x-\delta}^{-x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b - b') i h \ db' \ db,
\]

\[
A_{32}(x) = \int_{-x-\delta}^{-x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b - b') i h \ db' \ db,
\]

\[
A_{31}(x) = \int_{-x-\delta}^{-x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b - b') i h \ db' \ db,
\]

\[
A_{14}(x) = \int_{-y-\epsilon}^{-y+\epsilon} \int_{-\delta-x}^{-\delta-x+\delta} T(i, b' - b) i h \ db' \ db.
\]
\[ A_{31}(x) = \int_{x-\delta}^{x+\delta} \int_{y-\epsilon}^{y+\epsilon} T(h, b-b') ih' db' db, \]
and
\[ A_{23}(x) = \int_{-x-\delta}^{-x+\delta} \int_{y-\epsilon}^{y+\epsilon} T(h, b-b') h^2 db' db. \]

Now, the axiom requires that \( P(x + \Delta) - P(x) \geq 0 \). Equivalently, we require that \( \frac{P(x + \Delta) - P(x)}{2ih} \geq 0 \) for all \( h \), which implies in particular that
\[
\lim_{h \to 0} \frac{P(x + \Delta) - P(x)}{2ih} \geq 0.
\]

Using this information along with the computations for \( P(x) \) and the various \( A_{jk}(x) \)'s, we see (after some substitution of variables and transposition of terms) that
\[
\int_{-y-\epsilon}^{-y+\epsilon} \int_{x-\delta}^{x+\delta} [T(i, b' - b + \Delta) - T(i, b' - b)] db' db
\]
\[
\geq \int_{-y-\epsilon}^{-y+\epsilon} \int_{-x-\delta}^{-x+\delta} [T(i, b' - b) - T(i, b' - b - \Delta)] db' db.
\]

Dividing through by \( \delta \) in this expression and then taking \( \delta \) to zero, we may conclude that
\[
\int_{-y-\epsilon}^{-y+\epsilon} [T(i, x - b + \Delta) - T(i, x - b)] db
\]
\[
\geq \int_{-y-\epsilon}^{-y+\epsilon} [T(i, -x - b) - T(i, -x - b - \Delta)] db,
\]
and dividing this inequality, in turn, by \( \epsilon \) and taking \( \epsilon \) to zero, we see that
\[
T(i, a + \Delta) - T(i, a) \geq T(i, a') - T(i, a' - \Delta),
\]
where we use the observations that \( x + y = a \) and \( y - x = a' \). Therefore the conditions of Lemma 3 are satisfied, and \( T(i, \cdot) \) must be linear for every \( i > 0 \) since \( T(0, a) = 0 \). That is, there is a function \( \phi(i) \) such that \( T(i, a) = \phi(i) a \) for every \( i \) and \( a \). Given that \( T \) is continuous by assumption, the same must be true of \( \phi \).

Q.E.D.

**Lemma 5:** \( \phi(i) \) must be of the form \( Ki^\alpha \) for constants \( (K, \alpha) \gg 0 \).

**Proof:** As a preliminary step, observe that
\[
\phi(i) > 0 \quad \text{whenever} \quad i > 0; \tag{25}
\]
otherwise Axiom 3 would fail for configurations constructed from rectangular basic densities of equal height \( i \). We first prove that \( \phi \) satisfies the fundamental Cauchy equation
\[
\phi(p)\phi(p') = \phi(pp')\phi(1) \tag{26}
\]
for every \( (p, p') \gg 0 \). To this end, fix \( p \) and \( p' \) and define \( r \equiv pp' \). In what follows, we assume that \( p \geq r \).\(^{15}\) Consider a configuration with two basic densities, both of width \( 2\epsilon \), the first centered

\(^{15}\) If \( r \geq p \), simply permute \( p \) and \( r \) in the argument below.
at 0 and the second centered at 1. The heights are \( p \) and \( h \) (where \( h > 0 \) but soon to be made arbitrarily small). A little computation shows that polarization in this case is given by

\[
P = ph(\phi(p) + \phi(h))\left[\int_{-\epsilon}^{\epsilon} \int_{1-\epsilon}^{1+\epsilon} (b' - b) \, db' \, db\right]
\]

\[
+ \left[p^2 \phi(p) + h^2 \phi(h)\right] \left[\int_{-\epsilon}^{\epsilon} |b' - b| \, db' \, db\right]
\]

\[
= 4\epsilon^2 ph(\phi(p) + \phi(h)) + \frac{8\epsilon^3}{3}[p^2 \phi(p) + h^2 \phi(h)],
\]

where the first equality invokes Lemma 4. Now change the height of the first rectangle to \( r \).

Using (25) and \( p \geq r \), it is easy to see that for each \( \epsilon \), there exists a (unique) height \( h(\epsilon) \) for the second rectangle such that the polarizations of the two configurations are equated. Invoking (27), \( h(\epsilon) \) is such that

\[
ph(\phi(p) + \phi(h)) + \frac{2\epsilon}{3}[p^2 \phi(p) + h^2 \phi(h)]
\]

\[
= rh(\epsilon)[\phi(r) + \phi(h(\epsilon))] + \frac{2\epsilon}{3}[r^2 \phi(r) + h(\epsilon)^2 \phi(h(\epsilon))].
\]

By Axiom 4, it follows that for all \( \lambda > 0 \),

\[
\lambda^2 ph(\phi(\lambda p) + \phi(\lambda h)) + \frac{2\epsilon}{3}[(\lambda p)^2 \phi(\lambda p) + (\lambda h)^2 \phi(\lambda h)]
\]

\[
= \lambda^2 rh(\epsilon)[\phi(\lambda r) + \phi(h(\epsilon))] + \frac{2\epsilon}{3}[(\lambda r)^2 \phi(\lambda r) + (\lambda h(\epsilon))^2 \phi(h(\epsilon))].
\]

Notice that as \( \epsilon \downarrow 0 \), \( h(\epsilon) \) lies in some bounded set. We may therefore extract a convergent subsequence with limit \( h' \) as \( \epsilon \downarrow 0 \). By the continuity of \( \phi \), we may pass to the limit in (28) and (29) to conclude that

\[
ph(\phi(p) + \phi(h)) = rh'[\phi(r) + \phi(h')]
\]

and

\[
\lambda^2 ph(\phi(\lambda p) + \phi(\lambda h)) = \lambda^2 rh'[\phi(\lambda r) + \phi(\lambda h')].
\]

Combining (30) and (31), we see that

\[
\frac{\phi(p) + \phi(h)}{\phi(\lambda p) + \phi(\lambda h)} = \frac{\phi(r) + \phi(h')}{\phi(\lambda r) + \phi(\lambda h')}.\]

Taking limits in (32) as \( h \rightarrow 0 \) and noting that \( h' \rightarrow 0 \) as a result (examine (30) to confirm this), we have for all \( \lambda > 0 \),

\[
\frac{\phi(p)}{\phi(\lambda p)} = \frac{\phi(r)}{\phi(\lambda r)}.
\]

Put \( \lambda = 1/p \) and recall that \( r = pp' \). Then (33) yields the required Cauchy equation (26). To complete the proof, recall that \( \phi \) is continuous and that (25) holds. The class of solutions to (26) (that satisfy these additional qualifications) is completely described by \( \phi(p) = Kp^\alpha \) for constants \( (K, \alpha) \gg 0 \) (see, e.g., Aczél (1966, p. 41, Theorem 3)).

\[Q.E.D.\]

Lemmas 4 and 5 together establish “necessity,” though it still remains to establish the bounds on \( \alpha \). We shall do so along with our proof of “sufficiency,” which we begin now.
Lemma 6: Let $f$ be a basic density with mass $p$ and mean $\mu$ on support $[a, b]$. Let $m \equiv \mu - a$ and let $f^*$ denote the root of $f$. Then, if $f^\lambda$ denotes some $\lambda$-squeeze of $f$,

\begin{equation}
P(F^\lambda) = 4k p^{2+\alpha}(m\lambda)^{1-\alpha} \int_0^1 f^*(x)^{1+\alpha} \left\{ \int_0^1 f^*(y)(1-y) \, dy + \int_x^1 f^*(y)(y-x) \, dy \right\} \, dx
\end{equation}

for some constant $k > 0$.

Proof: Recall that a slide of $f$ has no effect on the computations, so we may as well set $a = 0$ and $b = 2m$, where $m = \mu - a$ is now to be interpreted as the mean. Given (3),

\begin{equation}
P(F) = k \int \int f(x)^{1+\alpha} f(y) |y - x| \, dy \, dx
\end{equation}

for some $k > 0$. Using the fact that $f$ is symmetric, we can write

\begin{equation}
P(F) = 2k \int_0^m \int_0^{2m} f(x')^{1+\alpha} f(y')(x' - y') \, dy' \, dx'
\end{equation}

\begin{align*}
&= 2k \int_0^m f(x')^{1+\alpha} \left\{ \int_0^{x'} f(y')(x' - y') \, dy' + \int_{x'}^m f(y')(y' - x') \, dy' \\
&\quad + \int_x^m f(y')(y' - x') \, dy' \right\} \, dx'.
\end{align*}

Examine the very last term in (36). Change variables by setting $z = 2m - y'$, and use symmetry to deduce that

\[
\int_{m}^{2m} f(y')(y' - x') \, dy' = \int_0^m f(z)(2m - x' - z) \, dz.
\]

Substituting this in (36), and manipulating terms, we obtain

\begin{equation}
P(F) = 4k \int_0^m f(x')^{1+\alpha} \left\{ \int_0^m f(y')(m - y') \, dy' + \int_{x'}^m f(y')(y' - x') \, dy' \right\} \, dx'.
\end{equation}

Now suppose that $f^\lambda$ is a $\lambda$-squeeze of $f$. Note that (37) holds just as readily for $f^\lambda$ as for $f$. Therefore, using the expression for $f$ given in (2), we see that

\[
P(F^\lambda) = 4k \lambda^{-(2+\alpha)} \int_{(1-\lambda)m}^m f\left( \frac{x' - (1-\lambda)m}{\lambda} \right)^{1+\alpha} \left\{ \int_{(1-\lambda)m}^m f\left( \frac{y' - (1-\lambda)m}{\lambda} \right)(m - y') \, dy' \\
+ \int_{x'}^m f\left( \frac{y' - (1-\lambda)m}{\lambda} \right)(y' - x') \, dy' \right\} \, dx'.
\]

Perform the change of variables

\[
x'' = \frac{x' - (1-\lambda)m}{\lambda} \quad \text{and} \quad y'' = \frac{y' - (1-\lambda)m}{\lambda}.
\]

Then it is easy to see that

\[
P(F^\lambda) = 4k \lambda^{1-\alpha} \int_0^m f(x'')^{1+\alpha} \left\{ \int_0^m f(y'')(m - y'') \, dy'' + \int_{x''}^m f(y'')(y'' - x'') \, dy'' \right\} \, dx''.
\]
To complete the proof, we must recover the root $f^*$ from $f$. To this end, first population-scale $f$ to $h$, where $h$ has mass 1. That is, $f(z) = ph(z)$ for all $z$. Doing so, we see that

$$P(F^*) = 4kp^{2+a}λ^{1-a} \int_0^m h(x')^{1+a} \left\{ \int_0^m h(y')(m-y') \,dy' \right\} \,dx'. $$

Finally, make the change of variables $x = x'/m$ and $y = y'/m$. Noting that $f^*(z) = mh(mz)$, we get (34).

**Q.E.D.**

**LEMMA 7:** Let $f$ and $g$ be two basic densities with disjoint support, with their means separated by distance $d$, and with population masses $p$ and $q$ respectively. Let $f$ have mean $\mu$ on support $[a, b]$. Let $m \equiv \mu - a$ and let $f^*$ denote the root of $f$. Then, for any $\lambda$-squeeze $f^\lambda$ of $f$,

$$A(f^\lambda, g) = 2kd)p^{1+\alpha}q(m\lambda)^{-\alpha} \int_0^1 f^*(x)^{1+\alpha} \,dx,$$

where $A(f^\lambda, g)$ denotes the total effective antagonism felt by members of $f^\lambda$ towards members of $g$.

**PROOF:** Without loss of generality, let $f$ have support $[0, 2m]$ (with mean $m$) and $g$ have support $[d, d + 2m]$ (where $d \geq 2m$ for disjoint supports). Using (35),

$$A(f, g) = k \int_0^{2m} f(x)^{1+\alpha} \left[ \int_d^{d+2m} g(y)(y-x) \,dy \right] \,dx$$

$$= k \int_0^{2m} f(x)^{1+\alpha} \left[ \int_d^{d+m} g(y)(y-x) \,dy + \int_{d+m}^{d+2m} g(y)(y-x) \,dy \right] \,dx$$

$$= k \int_0^{2m} f(x)^{1+\alpha} \left[ \int_d^{d+m} g(y)2(m+d-x) \,dy \right] \,dx$$

$$= kq \int_0^{2m} f(x)^{1+\alpha}(m+d-x) \,dx$$

$$= 2dkq \int_0^m f(x)^{1+\alpha} \,dx,$$

where the third equality exploits the symmetry of $g$,\(^1\)\(6\) the fourth equality uses the fact that $\int_d^{d+m} g(y) = q/2$, and the final equality uses the symmetry of $f$.\(^1\)\(7\) To be sure, this formula applies to any $\lambda$-squeeze of $f$, so that

$$A(f^\lambda, g) = 2dkq \int_0^m f^\lambda(x')^{1+\alpha} \,dx'$$

$$= 2dkq\lambda^{-\alpha} \int_{(1-\lambda)m}^{m} f\left(\frac{x'-(1-\lambda)m}{\lambda}\right)^{1+\alpha} \,dx'.$$

---

\(^{16}\)That is, for each $y \in [d, d + m]$, $g(y) = g(d + 2m - (y - d)) = g(2d + 2m - y)$. Moreover, $[y - x] + [(2d + 2m - y) - x] = 2(d + m - x)$.

\(^{17}\)That is, for each $x \in [0, m]$, $f(x) = f(2m - x)$. Moreover, $[m + d - x] + [m + d - (2m - x)] = 2d$. 
and making the change of variables

\[ x'' = x' - (1 - \lambda)m, \]

we may conclude that

\[ A(f^h, g) = 2dkq\lambda^{-\alpha} \int_0^m f(x'')^{1+\alpha} \, dx''. \]

To complete the proof, we must recover the root \( f^* \) from \( f \). As in the proof of Lemma 6, first population-scale \( f \) to \( h \), where \( h \) has mass 1. That is, \( f(z) = ph(z) \) for all \( z \). Doing so, we see that

\[ A(f^h, g) = 2dkp^{1+\alpha}q\lambda^{-\alpha} \int_0^m h(x'')^{1+\alpha} \, dx''. \]

Finally, make the change of variables \( x = x''/m \). Noting that \( f^*(z) = mh(mz) \), we get (38).

**Q.E.D.**

**Lemma 8:** Define, for any root \( f \) and \( \alpha > 0 \),

\[ \psi(f, \alpha) \equiv \frac{\int_0^1 f(x)^{1+\alpha} \, dx}{\int_0^1 f(x)^{1+\alpha}[\int_0^1 f(y)(1 - y) \, dy + \int_0^1 f(y)(y - x) \, dy] \, dx}. \]

Then—for any \( \alpha > 0 \)—\( \psi(f, \alpha) \) attains its minimum value when \( f \) is the uniform root, and this minimum value equals 3.

**Proof:** It will be useful to work with the inverse function

\[ \zeta(f, \alpha) \equiv \psi(f, \alpha)^{-1} = \frac{\int_0^1 f(x)^{1+\alpha}[\int_0^1 f(y)(1 - y) \, dy + \int_0^1 f(y)(y - x) \, dy] \, dx}{\int_0^1 f(x)^{1+\alpha} \, dx}. \]

Note that \( \zeta(f, \alpha) \) may be viewed as a weighted average of

\[ L(x) \equiv \int_0^1 f(y)(1 - y) \, dy + \int_x^1 f(y)(y - x) \, dy \]

as this expression varies over \( x \in [0, 1] \), where the “weight” on a particular \( x \) is just

\[ \frac{f(x)^{1+\alpha}}{\int_0^1 f(z)^{1+\alpha} \, dz}, \]

which integrates over \( x \) to 1. Now observe that \( L(x) \) is decreasing in \( x \). Moreover, by the unimodality of a root, the weights must be nondecreasing in \( x \). It follows that

\[ \zeta(f, \alpha) \leq \int_0^1 L(x) \, dx. \]

Now

\[ L(x) = \int_0^1 f(y)(1 - y) \, dy + \int_x^1 f(y)(y - x) \, dy = \int_0^1 f(y)(1 - x) \, dy + \int_0^x f(y)(x - y) \, dy = \frac{1 - x}{2} + \int_0^x f(y)(x - y) \, dy. \]
Because \( f(x) \) is nondecreasing and integrates to 1/2 on \([0, 1]\), it must be the case that
\[
\int_0^x f(y)(x - y)\,dy \leq \frac{\int_0^x (x - y)}{2} \,dy
\]
for all \( x \leq 1 \). Using this information in (42) and combining it with (41),
\[
(43) \quad \zeta(f, \alpha) \leq \int_0^1 \left[ \frac{1-x}{2} + \int_0^x \frac{y-x}{2} \,dy \right] \,dx
\]
\[
= \int_0^1 \left[ \int_0^1 \frac{1-y}{2} \,dy + \int_0^1 \frac{y-x}{2} \,dy \right] \,dx
\]
\[
= \zeta(u, \alpha),
\]
where \( u \) stands for the uniform root taking constant value 1/2 on \([0, 2]\). Simple integration reveals that
\( \zeta(u, \alpha) = 1/3 \). \( Q.E.D. \)

**Lemma 9:** Given that \( P(f) \) is of the form (35), Axiom 1 is satisfied if and only if \( \alpha \leq 1 \).

**Proof:** Simply inspect (34). \( Q.E.D. \)

**Lemma 10:** Given that \( P(f) \) is of the form (35), Axiom 2 is satisfied if and only if \( \alpha \geq 25 \).

**Proof:** Consider a configuration as given in Axiom 2: a symmetric distribution made out of three basic densities. By symmetry, the side densities must share the same root; call this \( f^* \). Let \( p \) denote their (common) population mass and \( m \) their (common) difference from their means to their lower support. Likewise, denote the root of the middle density by \( g^* \), by \( q \) its population mass, and by \( n \) the difference between mean and lower support. As in the proof of Lemma 2, we may decompose the polarization measure (35) into several components. First, there are the “internal polarizations” of the middle density \( P_m \) and of the two side densities \( P_s \). Next, there are various subtotals of effective antagonism felt by members of one of the basic densities towards another basic density. Let \( A_m \) denote this when the “origin” density is the middle and the “destination” density one of the sides. Likewise, \( A_m \) is obtained by permuting origin and destination densities. Finally, denote by \( A_s \) the total effective antagonism felt by inhabitants of one side towards the other side. Observe that each of these last four terms appear twice, so that (writing everything as a function of \( \lambda \)), overall polarization is given by
\[
P(\lambda) = P_m + 2P_s(\lambda) + 2A_m(\lambda) + 2A_m(\lambda) + 2A_s(\lambda)
\]
(44)
Compute these terms. For brevity, define for any root \( h \)
\[
\psi_1(h, \alpha) \equiv \int_0^1 h(x)^{1+\alpha} \left\{ \int_0^1 h(y)(1-y) \,dy + \int_x^1 h(y)(y-x) \,dy \right\} \,dx
\]
and
\[
\psi_2(h, \alpha) \equiv \int_0^1 h(x)^{1+\alpha} \,dx.
\]
Now, using Lemmas 6 and 7, we see that
\[
P_s(\lambda) = 4kp^{2+\alpha}(m\lambda)^{1-\alpha} \psi_1(f^*, \alpha),
\]
while
\[
A_m(\lambda) = 2kdq^{1+\alpha}pn^{-\alpha} \psi_2(g^*, \alpha).
\]
Moreover,
\[
A_m(\lambda) = 2kdq^{1+\alpha}q(m\lambda)^{-\alpha} \psi_2(f^*, \alpha)
\]
and
\[ A_{m}(\lambda) = 4kd\overline{p}^{2+\alpha}(m\lambda)^{-\alpha}\varphi_2(f^*, \alpha) \]
(where it should be remembered that the distance between the means of the two side densities is 2d). Observe from these calculations that \( A_{m}(\lambda) \) is entirely insensitive to \( \lambda \). Consequently, feeding all the computed terms into (44), we may conclude that
\[ P(\lambda) = C\left[2\lambda^{1-a} + \frac{d}{m}\psi(f^*, \alpha)\lambda^{-a}\left(\frac{q}{p} + 2\right)\right] + D, \]
where \( C \) and \( D \) are positive constants independent of \( \lambda \), and
\[ \psi(f^*, \alpha) = \frac{\varphi_2(f^*, \alpha)}{\varphi_1(f^*, \alpha)} \]
by construction; see (39) in the statement of Lemma 8. It follows from this expression that for Axiom 2 to hold, it is necessary and sufficient that for every three-density configuration of the sort described in that axiom,
\[ 2\lambda^{1-a} + \frac{d}{m}\psi(f^*, \alpha)\lambda^{-a}\left(\frac{q}{p} + 2\right) \]
must be nonincreasing in \( \lambda \) over \((0, 1]\). An examination of the expression in (45) quickly shows that a situation in which \( q \) is arbitrarily close to zero (relative to \( p \)) is a necessary and sufficient test case. By the same logic, one should make \( d/m \) as small as possible. The disjoint-support hypothesis of Axiom 2 tells us that this lowest value is 1. So it will be necessary and sufficient to show that for every root \( f^* \),
\[ \lambda^{1-a} + \psi(f^*, \alpha)\lambda^{-a} \]
is nonincreasing in \( \lambda \) over \((0, 1]\). For any \( f^* \), it is easy enough to compute the necessary and sufficient bounds on \( \alpha \). Simple differentiation reveals that
\[ (1 - \alpha)\lambda^{-a} - \alpha\psi(f^*, \alpha)\lambda^{-(1+a)} \]
must be nonnegative for every \( \lambda \in (0, 1] \); the necessary and sufficient condition for this is
\[ \alpha \geq \frac{1}{1 + \psi(f^*, \alpha)}. \]
Therefore, to find the necessary and sufficient bound on \( \alpha \) (uniform over all roots), we need to minimize \( \psi(f^*, \alpha) \) by choice of \( f^* \), subject to the condition that \( f^* \) be a root. By Lemma 8, this minimum value is 3. Using this information in (47), we are done. \( Q.E.D. \)

**Lemma 11:** Given that \( P(f) \) is of the form (35), Axiom 3 is satisfied.

**Proof:** Consider a symmetric distribution composed of four basic densities, as in the statement of Axiom 3. Number the densities 1, 2, 3, and 4, in the same order displayed in Figure 6. Let \( x \) denote the amount of the slide (experienced by the inner densities) in the axiom. For each such \( x \), let \( d_{jk}(x) \) denote the (absolute) difference between the means of basic densities \( j \) and \( k \). As we have done several times before, we may decompose the polarization of this configuration into several components. First, there is the "internal polarization" of each rectangle \( j \); call it \( P_j \), \( j = 1, 2, 3, 4 \). (These will stay unchanged with \( x \).) Next, there is the total effective antagonism felt by inhabitants of each basic density towards another; call this \( A_{jk}(x) \), where \( j \) is the "origin"
density and $k$ is the “destination” density. Thus total polarization $P(x)$, again written explicitly as a function of $x$, is given by

$$P(x) = \sum_{j=1}^{4} P_j + \sum_{j} \sum_{k \neq j} A_{jk}(x)$$

so that, using symmetry,

$$P(x) - P(0) = 2\left[ A_{12}(x) + A_{13}(x) - [A_{12}(0) + A_{13}(0)] + [A_{23}(x) - A_{23}(0)] \right]$$

Now Lemma 7 tells us that for all $i$ and $j$,

$$A_{ij}(x) = k_{ij}d_{ij}(x),$$

where $k_{ij}$ is a positive constant that is independent of distances across the two basic densities, and in particular is independent of $x$. Using this information in (48), it is trivial to see that

$$P(x) - P(0) = A_{23}(x) - A_{23}(0) = k_{ij}x > 0,$$

so that Axiom 3 is satisfied.

Q.E.D.

Given (35), Axiom 4 is trivial to verify. Therefore Lemmas 9, 10, and 11 complete the proof of the theorem.

Q.E.D.

REFERENCES


