A Group-Based Approach to Measuring Polarization

Supplementary Information

Isaac D. Mehlhaff*

October 19, 2022

Contents

*The University of North Carolina at Chapel Hill; mehlhaff@live.unc.edu.

S1 Analysis of Existing Literature

This section details the procedures and results of a simple exercise to gauge the use of various polarization measures in the political science literature. Journal articles were selected for the analysis if they satisfied two criteria: 1) Their title or abstract included any word with the root "polar," and 2) they were published between 2000 and 2021.[1] Polarization is an important concept throughout the social sciences but, for brevity and consistency, I focus on eight high-impact journals in political science. Four of these are general interest journals: *American Political Science Review*, *American Journal of Political Science*, *The Journal of Politics*, and *British Journal of Political Science*. The other four are subfield journals, selected to cover a wide range of subfields in which polarization is a prominent focus: *Public Opinion Quarterly*, *Legislative Studies Quarterly*, *Comparative Political Studies*, and *Political Analysis*. Table S1 describes the procedures used to obtain articles from each journal, including the database and search query used as well as any additional filters that needed to be applied to the results. These searches yielded a total of 322 articles.

All types of scholarship—empirical, theoretical, and formal theoretical—were included, but the vast majority included at least some quantitative empirical analysis. After obtaining these 322 articles, I read each of them to determine how the authors operationalized polarization.[2] In most cases, this information was contained in the "Data and Measures" section, or another section titled similarly.[3]

Before presenting the results, a few words are in order with regard to coding rules. When possible, I attempted to group similar operationalizations together. The "difference" category subsumed any operationalization measuring the distance between observations or group means, or extremity as distance from a scale midpoint. Any operationalization measuring the sum of distances from an overall or group mean or otherwise measuring the cohesion of a group was coded as "variance." Both of these measures are often adjusted, for example, with party vote shares. Many authors attempt to capture some sense of bimodality with their measurement, with kurtosis and the Reynal-Querol index being popular choices; these are included under "bimodality." The finally category garnering an appreciable number of articles is that of "overlap;"

[1] This excludes articles published online in "First View" or other similar pre-publication releases. That is, articles needed to be assigned an issue number to be included.
[2] All articles except for one specified this information in the main text. This article was coded as "unspecified."
[3] Eighty-two articles engaged with polarization theoretically but did not include it in any empirical analysis. These articles were coded as such and excluded from the calculations in Table S2.

Journal	Database	Search Query	Additional Filters	N
APSR	ProQuest	(ab(polariz*) AND PUBID(41041)) OR (ti(polariz*) AND PUBID(41041))	Date from 2000-01-01 to 2021-12-31	42
AJPS	JSTOR	polar* in Item Title field OR polar* in Abstract field	Date from 2000 to 2021, Journal is *American Journal of Political Science*	54
JOP	EBSCOhost Academic Search Premier	(TI(polar*) AND SO(”the journal of politics”)) OR (AB(polar*) AND SO(”the journal of politics”))	Date from 01-2000 to 12-2021, Journal is *Journal of Politics*	71
BJPS	ProQuest	(ab(polariz*) AND PUBID(48551)) OR (ti(polariz*) AND PUBID(48551))	Date from 2000-01-01 to 2021-12-31	41
POQ	EBSCOhost Academic Search Premier	(TI(polar*) AND SO(”public opinion quarterly”)) OR (AB(polar*) AND SO(”public opinion quarterly”))	Date from 01-2000 to 12-2021	39
LSQ	Wiley Online Library	polar* in Abstract field	Date from 1-2000 to 12-2021	35
CPS	SAGE Premier	polar* in Abstract field	Date from 2000 to 2021	27
PA	ProQuest	(ab(polariz*) AND PUBID(51675)) OR (ti(polariz*) AND PUBID(51675))	Date from 2000-01-01 to 2021-12-31	13

Table S1: Bibliometric Search Procedures

there are several measures purporting to measure the degree to which two distributions (generally displayed as kernel density plots) overlap, and these are all subsumed under this category. Other categories contain, at most, a few articles and are generally self-explanatory. As articles often employ multiple operationalizations, I allow articles to be coded with more than one measure when appropriate. Proportions therefore may not sum to 1.

Table S2 displays the proportion of articles in each journal and overall that use each measure of polarization. Three patterns in the results characterize the state of polarization measurement in the political science literature. First, there is wide variability in how scholars attempt to tap into polarization—this simple analysis reveals no less than twenty distinct measurement strategies. In one sense, then, there is little consistency in measurement. In contrast, two measures—difference and variance—are substantially more common than all others, appearing in 56.5% and 13.7% of all articles, respectively. In this sense, then, there is great consistency in measurement. However, the reliability of past findings may be called into question if these two measures do not adequately capture polarization, as I argue in the main text. Finally, there is some variation among subfields. In contrast to work in American politics, comparative scholarship tends to rely less on measures of difference and more on measures of variance, likely because the former complicates measurement in multi-party systems—a benefit to using the CPC that I show in the main text.

	APSR	*AJPS*	*JOP*	*BJPS*	*POQ*	*LSQ*	*CPS*	*PA*	**Overall**
Difference	0.548	0.63	0.62	0.512	0.538	0.629	0.333	0.615	**0.565**
Variance	0.19	0.13	0.099	0.195	0.077	0.057	0.259	0.154	**0.137**
Bimodality	0.048	0.037	0.014	0.049	0.026	0	0.074	0	**0.031**
Overlap	0	0.056	0	0.049	0.051	0.029	0	0.154	**0.031**
Correlation	0	0.037	0.014	0.098	0.026	0	0.037	0	**0.028**
Social Distance	0.071	0.019	0	0.024	0.077	0	0	0	**0.025**
Regression Coefficient	0	0	0.014	0.024	0.051	0.029	0	0	**0.016**
Proportion Extreme	0.024	0	0	0.024	0	0	0.074	0.077	**0.016**
Time	0	0	0	0.024	0	0.057	0	0	**0.009**
Party Unity	0	0.019	0	0	0	0.029	0	0	**0.006**
Party Votes	0	0	0.014	0	0	0	0.037	0	**0.006**
Divided Government	0	0	0	0.024	0	0	0.037	0	**0.006**
Coalition Size	0	0	0	0.024	0	0	0.037	0	**0.006**
Importance	0	0.019	0	0	0	0	0	0	**0.003**
R-Squared	0	0.019	0	0	0	0	0	0	**0.003**
Seat Proportion	0	0	0.014	0	0	0	0	0	**0.003**
Same-Party Clerk	0	0	0.014	0	0	0	0	0	**0.003**
Network Separation	0	0	0	0.024	0	0	0	0	**0.003**
Outparty Opinion	0	0	0	0	0.026	0	0	0	**0.003**
Engagement	0	0	0	0	0	0	0	0.077	**0.003**
Unspecified	0	0	0	0	0	0.029	0	0	**0.003**

Table S2: Proportion of Articles Using Polarization Measures

S2 Derivation and Properties of the Measure

S2.1 Derivation

To derive the CPC, I begin by decomposing the total variance of clustered data (TSS) in (S1) into components directly corresponding to the two features of polarization: the variance accounted for between the clusters (BSS, corresponding to intergroup heterogeneity) and the variance accounted for within all clusters (WSS, corresponding to intragroup homogeneity). Dividing by TSS and solving for the BSS term gives an expression for the proportion of the total variance accounted for by the between-cluster variance—what I call the cluster-polarization coefficient (CPC):

$$\begin{aligned} TSS &= BSS + WSS, \\ \rightarrow CPC &= 1 - \frac{WSS}{TSS} = \frac{BSS}{TSS}. \end{aligned} \tag{S1}$$

More formally, I compute three terms in (S2): the total sum of squares TSS, the between-cluster sum of squares BSS, and the total within-cluster sum of squares WSS, where each individual i in cluster k holds a position on dimension j:

$$
\begin{aligned}
TSS &= \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{ij} - \mu_j)^2, \\
BSS &= \sum_{k=1}^{n_k} \sum_{j=1}^{n_j} (\mu_{kj} - \mu_j)^2, \\
WSS &= \sum_{k=1}^{n_k} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{ikj} - \mu_{kj})^2.
\end{aligned}
\tag{S2}
$$

Mirroring (S1), I arrive at formal expressions of the variance of clustered data and of the CPC in (S3). Expressed in this way, the CPC appears related—though not identical—to a one-way ANOVA F-statistic and the coefficient of determination (R^2).

$$
\begin{aligned}
TSS &= BSS + WSS, \\
\rightarrow \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{ij} - \mu_j)^2 &= \sum_{k=1}^{n_k} \sum_{j=1}^{n_j} (\mu_{kj} - \mu_j)^2 + \sum_{k=1}^{n_k} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{ikj} - \mu_{kj})^2, \\
\rightarrow CPC &= 1 - \frac{\sum_{k=1}^{n_k} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{ikj} - \mu_{kj})^2}{\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{ij} - \mu_j)^2}, \\
&= \frac{\sum_{k=1}^{n_k} \sum_{j=1}^{n_j} (\mu_{kj} - \mu_j)^2}{\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{ij} - \mu_j)^2}, \\
&\forall \quad i \in (1, ..., n_i), \quad j \in (1, ..., n_j), \quad k \in (1, ..., n_k).
\end{aligned}
\tag{S3}
$$

The CPC thus possesses two desirable properties: It is naturally bounded on the interval $[0, 1]$ and it takes into account both features of polarization. The CPC increases when the distance between groups increases or when groups become more tightly concentrated around their collective ideal point, but the rate of those increases depends on the relative levels of BSS and WSS.

As I show in section S2.3, however, this measure will be biased upward in small samples. To make the CPC more generalizable to contexts with varying numbers of observations, variables, and clusters, I incorporate corrections for lost degrees of freedom into the key variance expressions in (S4) and derive the adjusted CPC in (S5):

$$
\begin{aligned}
TSS_{adj} &= \frac{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}(x_{ij}-\mu_j)^2}{n_i - n_j}, \\
WSS_{adj} &= \frac{\sum_{k=1}^{n_k}\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}(x_{ikj}-\mu_{kj})^2}{n_i - n_j n_k}.
\end{aligned} \tag{S4}
$$

$$
\begin{aligned}
CPC_{adj} &= 1 - \frac{WSS_{adj}}{TSS_{adj}}, \\
\rightarrow CPC_{adj} &= 1 - \frac{\frac{\sum_{k=1}^{n_k}\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}(x_{ikj}-\mu_{kj})^2}{n_i - n_j n_k}}{\frac{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}(x_{ij}-\mu_j)^2}{n_i - n_j}}, \\
&= 1 - \frac{\sum_{k=1}^{n_k}\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}(x_{ikj}-\mu_{kj})^2}{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}(x_{ij}-\mu_j)^2}\frac{n_i - n_j}{n_i - n_j n_k}, \\
&= 1 - (1 - CPC)\frac{n_i - n_j}{n_i - n_j n_k}, \\
&\forall \quad i \in (1, ..., n_i), \quad j \in (1, ..., n_j), \quad k \in (1, ..., n_k).
\end{aligned} \tag{S5}
$$

S2.2 Distribution

Because the CPC is effectively a ratio of two variances, an F statistic can be calculated:[4]

[4]This set-up and the proofs which follow are similar to the derivation of a sampling distribution for R^2 (Magee 1990).

$$
\begin{aligned}
F &= \frac{\frac{BSS}{n_j n_k - n_j}}{\frac{WSS}{n_i - n_j n_k}}, \\
&= \frac{n_i - n_j n_k}{n_j n_k - n_j} \frac{TSS - WSS}{WSS}, \\
&= \frac{n_i - n_j n_k}{n_j n_k - n_j} \frac{TSS}{TSS} \frac{TSS - WSS}{WSS}, \\
&= \frac{n_i - n_j n_k}{n_j n_k - n_j} \frac{1 - \frac{WSS}{TSS}}{\frac{WSS}{TSS}}, \\
&= \frac{n_i - n_j n_k}{n_j n_k - n_j} \frac{1 - \frac{WSS}{TSS}}{1 - (1 - \frac{WSS}{TSS})}, \\
&= \frac{n_i - n_j n_k}{n_j n_k - n_j} \frac{CPC}{1 - CPC}.
\end{aligned} \tag{S6}
$$

where n_i, n_j, and n_k denote the number of observations, dimensions, and clusters, respectively; and $BSS \sim \chi^2(n_j n_k - n_j)$ and $WSS \sim \chi^2(n_i - n_j n_k)$ under the null hypothesis that $BSS - WSS = 0$. F is therefore distributed as a central $F(n_j n_k - n_j, n_i - n_j n_k)$ random variable. Solving for the CPC term from (S6) produces:

$$
\begin{aligned}
F &= \frac{n_i - n_j n_k}{n_j n_k - n_j} \frac{CPC}{1 - CPC}, \\
\rightarrow (n_j n_k - n_j)F - (n_j n_k - n_j)(CPC)F &= (n_i - n_j n_k)CPC, \\
\rightarrow (n_j n_k - n_j)F &= [(n_i - n_j n_k) + (n_j n_k - n_j)F]CPC, \\
\rightarrow CPC &= \frac{(n_j n_k - n_j)F}{(n_i - n_j n_k) + (n_j n_k - n_j)F},
\end{aligned} \tag{S7}
$$

implying that $CPC \sim \text{Beta}(\frac{n_j n_k - n_j}{2}, \frac{n_i - n_j n_k}{2})$ under the null—a sensible result given that the CPC and the Beta distribution both have continuous support on the [0, 1] interval.

S2.3 Unbiasedness

From this distribution, it is straightforward to recover the mean:

$$\begin{aligned} E(CPC) &= \frac{\frac{n_j n_k - n_j}{2}}{\frac{n_j n_k - n_j}{2} + \frac{n_i - n_j n_k}{2}}, \\ &= \frac{n_j n_k - n_j}{n_i - n_j}. \end{aligned} \tag{S8}$$

The unadjusted CPC is therefore asymptotically unbiased, again under the null hypothesis that $BSS - WSS = 0$, as $\lim_{n_i \to \infty} E(CPC) = 0$. Taking advantage of the expression in (S8), we can show that the adjusted CPC is an unbiased estimator under the null:

$$\begin{aligned} E(CPC_{adj}) &= E[1 - (1 - CPC)\frac{n_i - n_j}{n_i - n_j n_k}], \\ &= 1 - E[(1 - CPC)\frac{n_i - n_j}{n_i - n_j n_k}], \\ &= 1 - [E(1 - CPC)]\frac{n_i - n_j}{n_i - n_j n_k}, \\ &= 1 - [1 - E(CPC)]\frac{n_i - n_j}{n_i - n_j n_k}, \\ &= 1 - (1 - \frac{n_j n_k - n_j}{n_i - n_j})\frac{n_i - n_j}{n_i - n_j n_k}, \\ &= 1 - (\frac{n_i - n_j}{n_i - n_j} - \frac{n_j n_k - n_j}{n_i - n_j})\frac{n_i - n_j}{n_i - n_j n_k}, \\ &= 1 - (\frac{n_i - n_j n_k}{n_i - n_j})\frac{n_i - n_j}{n_i - n_j n_k}, \\ &= 1 - 1, \\ &= 0. \end{aligned} \tag{S9}$$

S2.4 Consistency

The variance of the CPC can also be recovered from its sampling distribution:

$$\begin{aligned}\text{Var}(CPC) &= \frac{\frac{n_j n_k - n_j}{2} \frac{n_i - n_j n_k}{2}}{(\frac{n_j n_k - n_j}{2} + \frac{n_i - n_j n_k}{2})^2 (\frac{n_j n_k - n_j}{2} + \frac{n_i - n_j n_k}{2} + 1)}, \\ &= \frac{2(n_j n_k - n_j)(n_i - n_j n_k)}{(n_i - n_j)^2 (n_i - n_j + 1)},\end{aligned} \tag{S10}$$

The unadjusted CPC is therefore consistent under the null, as $\lim_{n_i \to \infty} \text{Var}(CPC) = 0$. Taking advantage of (S10), we can also derive the variance of the adjusted CPC:

$$\begin{aligned}\text{Var}(CPC_{adj}) &= \text{Var}[1 - (1 - CPC)\frac{n_i - n_j}{n_i - n_j n_k}], \\ &= \text{Var}[(1 - CPC)\frac{n_i - n_j}{n_i - n_j n_k}], \\ &= (\frac{n_i - n_j}{n_i - n_j n_k})^2 \text{Var}(1 - CPC), \\ &= (\frac{n_i - n_j}{n_i - n_j n_k})^2 \text{Var}(CPC), \\ &= (\frac{n_i - n_j}{n_i - n_j n_k})^2 \frac{2(n_j n_k - n_j)(n_i - n_j n_k)}{(n_i - n_j)^2 (n_i - n_j + 1)}, \\ &= \frac{2(n_j n_k - n_j)}{(n_i - n_j n_k)(n_i - n_j + 1)}.\end{aligned} \tag{S11}$$

The adjusted CPC is therefore consistent under the null, as $\lim_{n_i \to \infty} \text{Var}(CPC_{adj}) = 0$.

S2.5 Additional Properties

Finally, I derive expressions for the median and mode of the sampling distribution. The median is presented in (S12), where I_x gives the regularized incomplete Beta function:[5]

[5] The solution given in (S12) is approximate—the median of the Beta distribution has no closed-form expression for arbitrary parameters.

$$\begin{aligned}\text{median}(CPC) &= I^{-1}_{\frac{1}{2}}\left(\frac{n_j n_k - n_j}{2}, \frac{n_i - n_j n_k}{2}\right), \\ &\approx \frac{\frac{n_j n_k - n_j}{2} - \frac{1}{3}}{\frac{n_j n_k - n_j}{2} + \frac{n_i - n_j n_k}{2} - \frac{2}{3}}, \\ &\approx \frac{3(n_j n_k - n_j) - 2}{3(n_i - n_j) - 4}.\end{aligned} \tag{S12}$$

The mode of the sampling distribution can be expressed as:

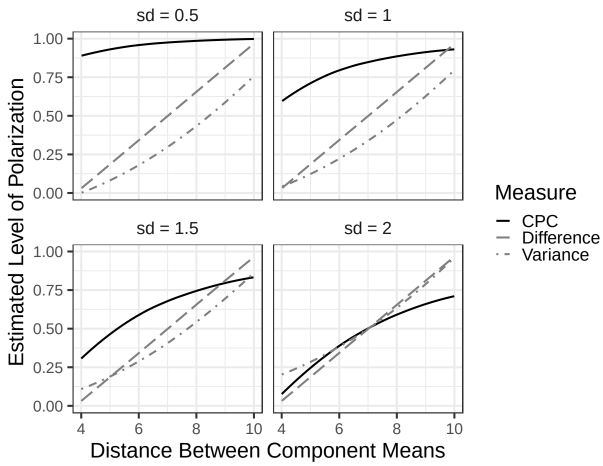
$$\begin{aligned}\text{mode}(CPC) &= \frac{\frac{n_j n_k - n_j}{2} - 1}{\frac{n_j n_k - n_j}{2} + \frac{n_i - n_j n_k}{2} - 2}, \\ &= \frac{n_j n_k - n_j - 2}{n_i - n_j - 4},\end{aligned} \tag{S13}$$

implying that the distribution possesses a unique and finite mode when $n_i - n_j > 4$ and $n_j n_k - n_j \geq 2$.

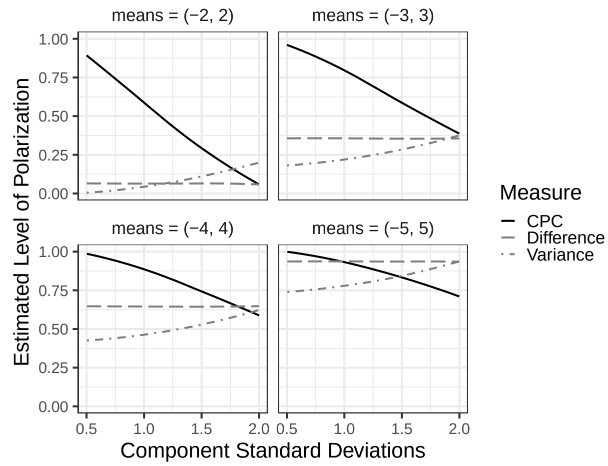
S3 Simulation Evidence

S3.1 Plots of Results with Three and Four Components

This section presents additional plots of simulation results, analogous to main text Figures 3 and 4, for the cases of three and four components.

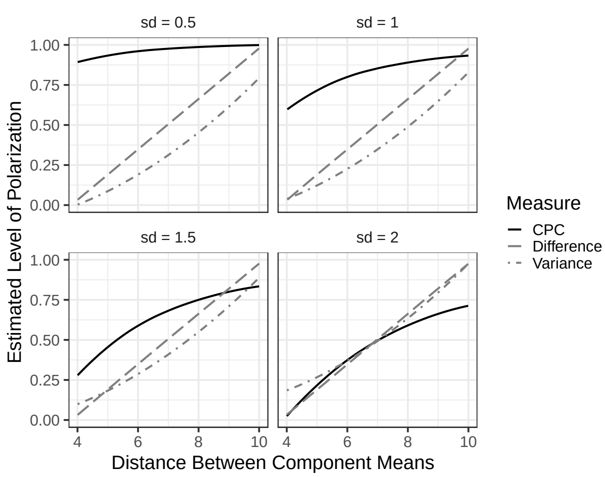


(a) Random μ (intergroup heterogeneity)

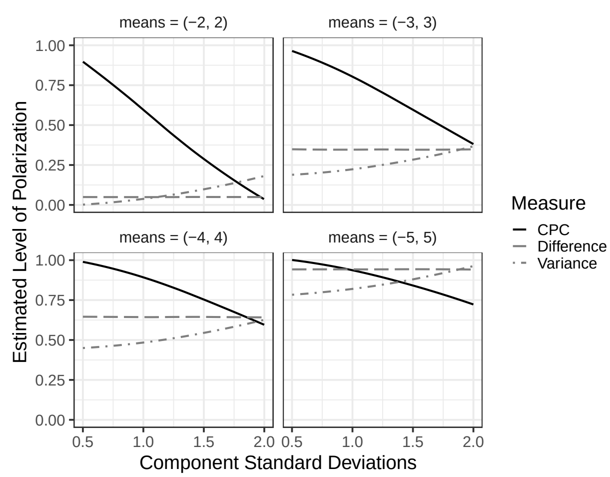


(b) Random σ (intragroup homogeneity)

Figure S1: Univariate Polarization Estimates with Three Components Results from univariate simulations of polarization measures with three components, showing estimated level of polarization for a randomly varying distribution parameter, holding the other parameter constant. All measures scaled to $[0, 1]$ to enable comparison and plotted using LOESS.

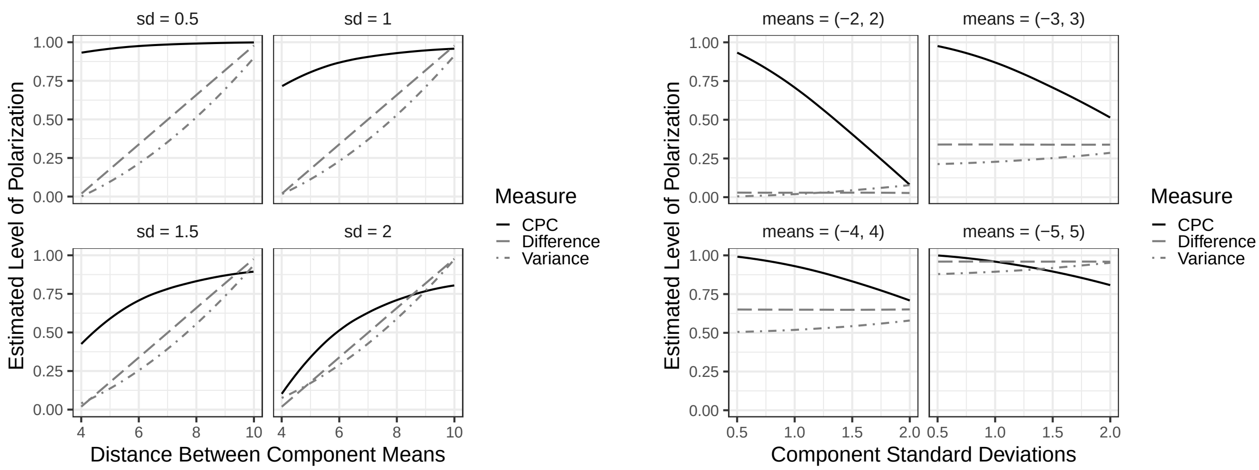


(a) Random μ (intergroup heterogeneity)



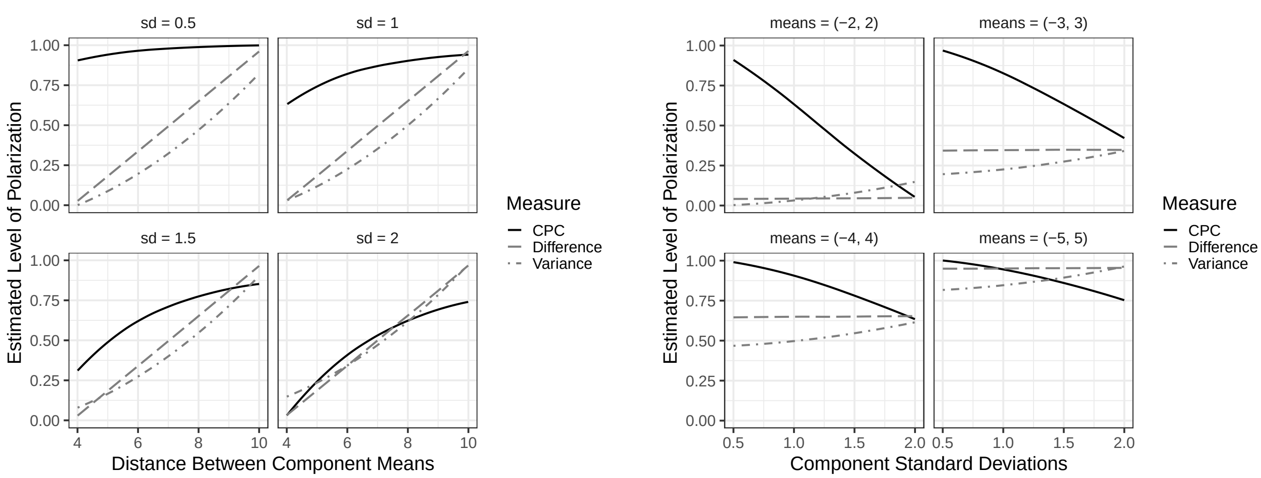
(b) Random σ (intragroup homogeneity)

Figure S2: Bivariate Polarization Estimates with Three Components. Results from bivariate simulations of polarization measures with three components, showing estimated level of polarization for a randomly varying distribution parameter, holding the other parameter constant. All measures scaled to $[0, 1]$ to enable comparison and plotted using LOESS.



(a) Random μ (intergroup heterogeneity) (b) Random σ (intragroup homogeneity)

Figure S3: Univariate Polarization Estimates with Four Components. Results from univariate simulations of polarization measures with four components, showing estimated level of polarization for a randomly varying distribution parameter, holding the other parameter constant. All measures scaled to $[0, 1]$ to enable comparison and plotted using LOESS.



(a) Random μ (intergroup heterogeneity) (b) Random σ (intragroup homogeneity)

Figure S4: Bivariate Polarization Estimates with Four Components. Results from univariate simulations of polarization measures with four components, showing estimated level of polarization for a randomly varying distribution parameter, holding the other parameter constant. All measures scaled to $[0, 1]$ to enable comparison and plotted using LOESS.

S3.2 Results with Log-Normal Data

Constructing the CPC using sums of squares imposes an important limitation: It is sensitive to extreme outliers. Consider how this measure would respond to adding observations far out in the tails of a distribution.

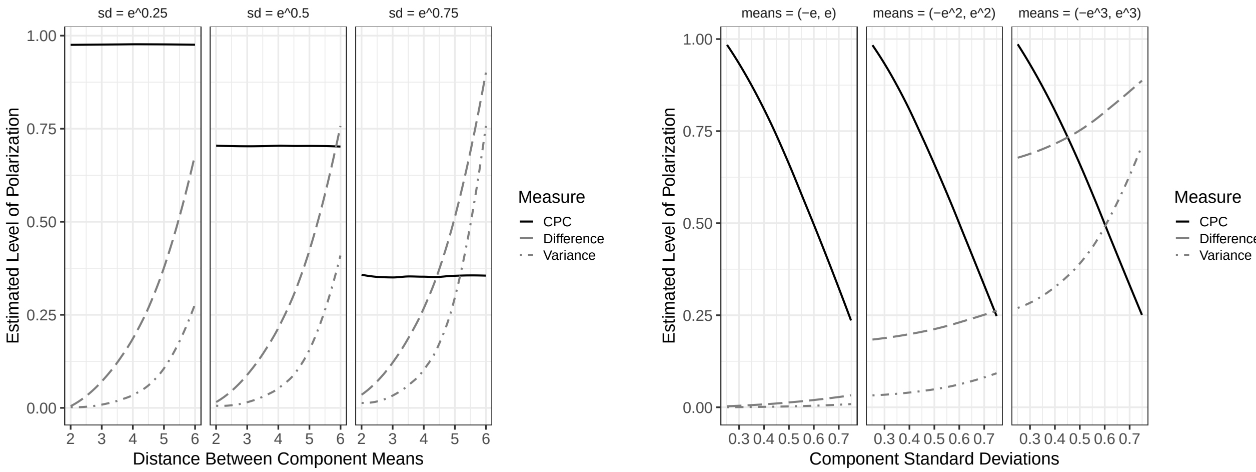
Such observations would lead to rapid increases in WSS, likely overwhelming any increase in BSS that would result from small changes to the group's centroid. As a consequence, the CPC will rapidly decrease and remain relatively insensitive to any further changes in either variable, for the same discussed in the main text. To demonstrate this limitation, I followed a simulation procedure similar to the one in the main text. Simulations to evaluate polarization measures on the intergroup heterogeneity feature follow the procedure below:

1. Fix component standard deviations at a range of values $\sigma \in \{e^{0.25}, e^{0.5}, e^{0.75}\}$. For identification, I use the same σ for each component.
2. For each component standard deviation σ, select 500 values of μ as independent draws from $\text{U}(e, e^3)$.
3. Take 500 independent draws from $\text{logN}(\mu, \sigma)$.
4. Reflect each data point over the y-axis to obtain a bimodal distribution with heavy tails on left and right.
5. Apply each polarization measure to the resulting distribution.

The result of this procedure is 1,000 heavy-tailed distributions, each with $N = 1000$. Simulations to evaluate polarization measures on the intragroup homogeneity feature follow the procedure below:

1. Fix component means at a range of values $\mu \in \{e, e^2, e^3\}$. For identification, I use the same absolute value of μ for each component.
2. For each component mean μ, select 500 values of σ as independent draws from $\text{U}(e^{0.25}, e^{0.75})$.
3. Take 500 independent draws from $\text{logN}(\mu, \sigma)$.
4. Reflect each data point over the y-axis to obtain a bimodal distribution with heavy tails on left and right.
5. Apply each polarization measure to the resulting distribution.

The result of this procedure is 1,000 heavy-tailed distributions, each with $N = 1000$. Figure S5 displays the results. As postulated, the CPC remains sensitive to component standard deviations (seen in plot (b)), but the rapid increase in WSS as a result of the heavy-tailed data overwhelms an increase in BSS and makes the measure insensitive to component means (seen in plot (a)). The performance of the other measures is similarly mixed. Difference and variance perform closer to expectations on the intergroup heterogeneity

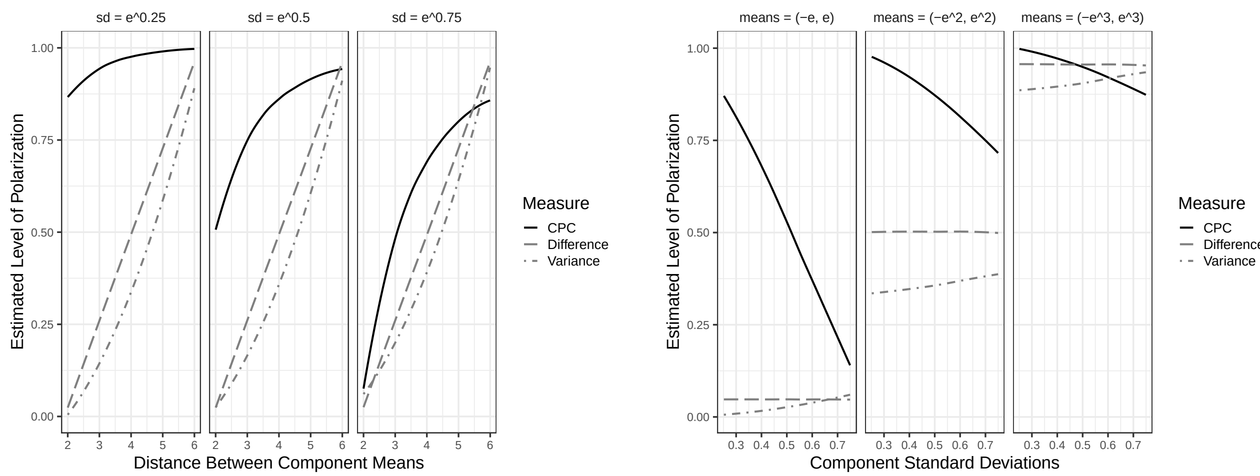


(a) Varying μ (intergroup heterogeneity) (b) Varying σ (intragroup homogeneity)

Figure S5: Polarization Estimates with Log-Normal Data. Results from univariate simulations of polarization measures with two components, showing estimated level of polarization for a randomly varying distribution parameter, holding the other parameter constant. All measures scaled to $[0, 1]$ to enable comparison and plotted using LOESS.

feature (plot a), but both extant measures move in the wrong direction in response to varying standard deviations (plot b).

These problems can be partially ameliorated, however, by taking the natural log of the data (between steps 3 and 4 of the simulation procedures). While this transformation maintains the heavy-tailed nature of the simulated distribution, it maps it to an interval that decreases the absolute value of the outlier observations and makes it more manageable for the sums of squares used to calculate the CPC. Figure S6 displays the results of applying each measure to the transformed data. The CPC displays behavior much more in line with expectations on both features, and the other measures generally improve as well, although they all still struggle to capture the intragroup homogeneity feature (plot (b)). This underscores the need for analysts to be aware of the distribution of their data. Although distributions of real-world data may not often display heavy-tailed behavior to the extent that this simulated data does, the measure can still be vulnerable to extreme outliers. Researchers should apply appropriate transformations to their data before estimating polarization, just as they would before fitting a linear model or conducting other analyses.



(a) Varying μ (intergroup heterogeneity)

(b) Varying σ (intragroup homogeneity)

Figure S6: Polarization Estimates with Transformed Log-Normal Data. Results from univariate simulations of polarization measures with two components, showing estimated level of polarization for a randomly varying distribution parameter, holding the other parameter constant and transforming the log-normal data using the natural log. All measures scaled to $[0, 1]$ to enable comparison and plotted using LOESS.

S3.3 Results with Unequal Component Weights

I noted in the main text that I maintain equal component weights in the simulations to aid in identification. This could raise concerns that results are at least partially dependent on an assumption that may not be reasonable in many real-world applications. To evaluate the effects of this constraint, I use a similar simulation framework as in the main text, but hold μ and σ constant and instead randomize cluster weights. Simulations to evaluate polarization measures on the intergroup heterogeneity feature follow the procedure below:

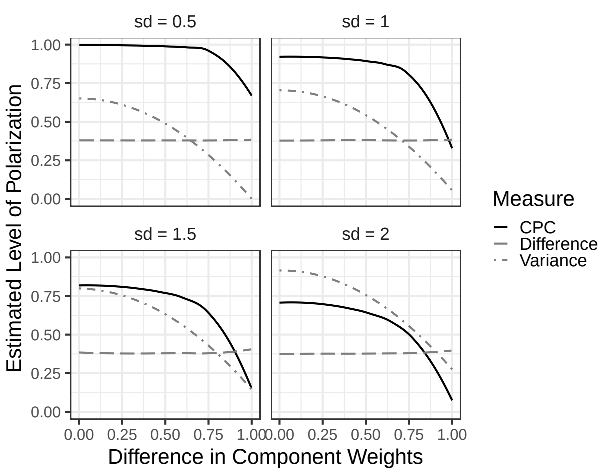
1. Fix component means at $(-3, 3)$, a middling level of polarization in this simulation framework.
2. Fix component standard deviations at a range of values $\sigma \in \{0.5, 1, 1.5, 2\}$. For identification, I use the same σ for each component and maintain a global mean of zero.
3. For each component, draw a set of component weights ϕ_K as independent draws from $\text{U}(0, 1)$, such that $\sum_{k=1}^{2} \phi_k = 1$.
4. Take 1,000 independent draws from a Gaussian mixture parameterized by $\text{N}(\phi_1, -\mu, \sigma; \phi_2, \mu, \sigma)$.
5. Apply each polarization measure to the resulting distribution.

The result of this procedure is 1,000 distributions, each with $N = 1000$. Simulations to evaluate polarization measures on the intragroup homogeneity feature follow the procedure below:

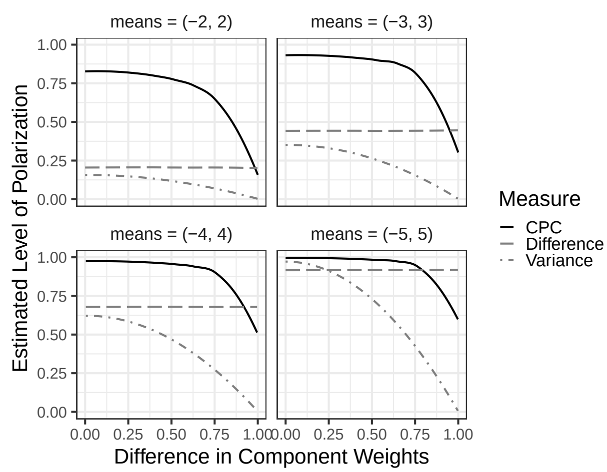
1. Fix component standard deviations at 1, a middling level of polarization in this simulation framework.
2. Fix component means at a range of values $\mu \in \{2, 3, 4, 5\}$. For identification, I use the same absolute value of μ for each component and maintain a global mean of zero.
3. For each component, draw a set of component weights ϕ_K as independent draws from $\text{U}(0, 1)$, such that $\sum_{k=1}^{2} \phi_k = 1$.
4. Take 1,000 independent draws from a Gaussian mixture parameterized by $\text{N}(\phi_1, -\mu, \sigma; \phi_2, \mu, \sigma)$.
5. Apply each polarization measure to the resulting distribution.

The result of this procedure is 1,000 distributions, each with $N = 1000$.

For each set of simulations, I then plot estimated polarization as a function of the difference between the two component weights. As in all other simulations, I calculate the CPC using the true cluster memberships, which are known from the data randomization procedure. Figure S7 displays the results. As expected, the adjusted CPC is relatively invariant to the difference in component weights except when that difference is high, at which point the mixture distribution approaches unimodality and the CPC decreases precipitously.



(a) Varying μ (intergroup heterogeneity)



(b) Varying σ (intragroup homogeneity)

Figure S7: Univariate Polarization Estimates with Two Components. Results from univariate simulations of polarization measures with two components, showing estimated level of polarization for randomly varying component weights, holding distribution parameters constant. All measures scaled to $[0, 1]$ to enable comparison and plotted using LOESS.

S4 Benchmarking Against Human Coders

The simulation evidence presented in the previous section is helpful for ensuring the CPC responds in theoretically appropriate ways to changes in distributional features. However, the lack of ground-truth polarization labels makes it difficult to judge whether the CPC is a more accurate measure of polarization compared to others. Recovering a ranking of distributions based on randomized distribution parameters comes close to solving that problem, but in these simulations, only one feature can be manipulated at a time. In real-world data, both features vary simultaneously, often independent of each other. I therefore use human coders to gather ground-truth annotations of distributions' level of polarization, against which I can benchmark each measure's performance.

I first generated fifty bimodal distributions using the same parameters as in section S4.1, randomly varying both component means and standard deviations at the same time. To make the task more accessible to non-experts, I colored one component blue and one component red and referred to them as graphs of Democratic and Republican ideology. The full task preamble and sample graphs can be seen in Figure S8.

Lots of people have noticed that American politics have become more polarized in recent years. Polarization means that Democrats have ideologies that are very similar to other Democrats' ideologies and very different from Republicans' ideologies, and vice versa.

We need to collect data about what this polarization looks like in graphs of Democrats' and Republicans' ideologies. We will show you 20 pairs of graphs like the ones below. In this pair, the graph on the right (or on the bottom if viewing on a mobile device) is more polarized because Democrats' (blue) and Republicans' (red) ideologies are farther apart and each party's members have ideologies that are more similar to others in their party. Your task will be to tell us which of the two graphs looks more polarized in each of the 20 pairs.

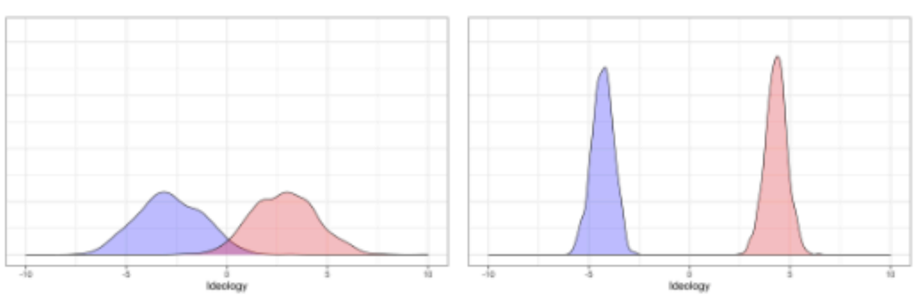


Figure S8: Preamble and Sample Graphs of Polarization Labeling Task

After the preamble, coders answered three screening questions to ensure they understood the task. These screening questions were identical to those they saw in the main task, but I chose the graphs presented in the screening questions such that one was obviously more polarized than the other. Coders were only allowed to complete the main task if they answered all three screening questions correctly. I explicitly told coders that polarization involved both features and tested them on that knowledge with easy screening questions for two reasons. First, I wanted to make sure coders understood the task. Second, I was not concerned with how the coders themselves thought about polarization, only that they could evaluate polarization according to the two-feature definition while *both* features varied across distributions—something I could not achieve in the simulations. For those coders who passed the screening questions, I randomly selected twenty pairs of plots, presented them to the coders, and asked them: "Which of the two graphs below is more polarized?"

The result of each coder's task was therefore twenty random pairwise comparisons labeled according to which of the two options was more polarized. 495 workers on Amazon's Mechanical Turk completed the task.[6] From the pairwise comparisons, a log-linear Bradley-Terry model recovered a ranking of distributions according to their level of polarization (Bradley and Terry 1952). I then applied each measure of polarization to all distributions and again ranked them according to the level of estimated polarization. The outcome of this exercise, then, is four rankings of fifty distributions: one set of rankings uncovered by measures of difference, variance, and the CPC, and one set of rankings recovered from human coding to be treated as ground-truth information.

	Difference	Variance	CPC
RMSE	18.17	20.381	**4.441**
MAE	14.88	16.88	**3.52**

Table S3: Benchmark of Polarization Measurements Against Rankings Recovered from Human Coders. Root mean squared error and mean absolute error; bolded values denote measure with lowest error.

Table S3 describes the extent to which each polarization measure departs from the human-coded ground-truth ranking. On both root mean squared error (RMSE) and mean absolute error (MAE), the CPC gives a more accurate approximation of the human-coded ranking, with error metrics less than one third the size of difference, the next most accurate measure. The magnitude of the difference and variance measures' error is also noteworthy. In this context, an MAE of 15 to 17 (on a 50-point scale) suggests that the difference between any given ranking and the true ranking will be, on average, approximately one third of the scale.

S4.1 Data Collection and Ethics

The human-coded data used in this section was collected in accordance with the American Political Science Association's standards for professional ethics and principles for human subjects research. Data

[6] In addition to passing the three screening questions, workers were required to be over age 18, located in the United States, and have at least a 95% approval record on more than fifty completed MTurk tasks.

collection and handling protocols were approved by an Institutional Review Board. The data collection task was programmed in Qualtrics and distributed to workers recruited through Amazon's Mechanical Turk crowdsourcing platform. Respondents are all Americans aged 18 years or older. Vulnerable populations (e.g. prisoners, respondents with a direct-dependency relationship with the researchers, decisionally impaired individuals, etc.) were unlikely to be included as respondents.

Prior to beginning the study, all respondents were informed of potential risks (consequences of breach of confidentiality) in a process of informed consent. They were also informed that they had the right to refuse the study, opt out of the study at any time, or withdraw their data after completing the survey. The task was approximately two minutes in length. Workers were paid $0.35 for the task, making worker compensation similar to what they could expect to earn for an equivalent amount of work at a minimum-wage job in most US cities.

The data set contains no personal identifiers. A randomly generated identifier with no relation to actual identifiers was generated by the Qualtrics software and used to connect each completed survey response to the Worker ID (a random alphanumeric string assigned by Amazon) of the individual who completed it. No demographic data was collected, so deductive disclosure was not possible.

S5 Cross-National Polarization Application

S5.1 Selecting Number of Clusters

As noted in the main text, one challenge in using clustering methods is the need to specify the number of clusters n_k. Although Figure 6 in the main text makes this selection relatively clear, I use silhouette scores to confirm the appropriate number of clusters in each distribution. This method involves repeatedly running the clustering algorithm with an increasingly large number of clusters. For each n_k, a "silhouette score" is then calculated. Silhouette scores use an arbitrary distance metric to determine how similar each point is to its own cluster and how different it is from other clusters. These individual scores are then aggregated into a value in the range $[-1, 1]$, where -1 implies the worst possible fit and 1 implies the best possible fit. The number of clusters which produces the maximum silhouette score is the most appropriate value for n_k. Figure S9 displays silhouette plots for each country. There is a clear maximum value for each country, corresponding to a two-cluster specification in Germany, Italy, the United States, and the United Kingdom, and three-cluster specifications in the Netherlands and Spain.

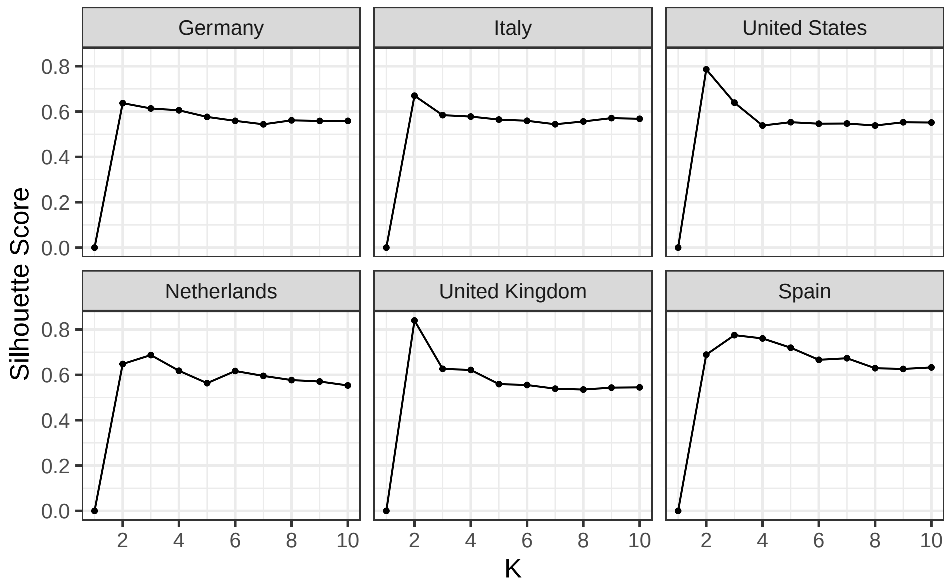


Figure S9: Silhouette scores of Twitter ideology estimates for each country.

S5.2 Full Polarization Estimates

Table S4 displays the polarization estimates and associated standard errors for each country and measure presented in the analysis of elite ideological ideal points in the main text. The estimates in Table S4 correspond to Figure 7 in the main text.

	Measure		
	Difference	Variance	CPC
Germany	2.6	1.93	0.697
	(0.118)	(0.121)	(0.023)
Italy	4.24	5.05	0.758
	(0.19)	(0.409)	(0.021)
Netherlands	3.45	4.0	0.912
	(0.184)	(0.372)	(0.013)
Spain	2.9	2.3	0.94
	(0.088)	(0.222)	(0.01)
United Kingdom	6.35	9.69	0.921
	(0.166)	(0.775)	(0.016)
United States	2.05	1.17	0.885
	(0.041)	(0.044)	(0.01)

Table S4: Polarization Estimates and Associated Standard Errors from Analysis of Cross-National Elite Ideal Points. Corresponds to main text Figure 7. Standard errors are in parentheses

S6 Affective Polarization Application

S6.1 Full Polarization Estimates

Table S5 displays polarization estimates and associated standard errors for all 28 countries included in the analysis of affective polarization in the main text. The estimates in Table S5 correspond to Figure 8 in the main text.

	Measure				*Measure*		
	Difference	Variance	CPC		Difference	Variance	CPC
Argentina	4.896	21.259	0.634	Montenegro	4.251	26.888	0.107
	(0.408)	(1.023)	(0.032)		(0.318)	(1.363)	(0.02)
Australia	5.071	22.755	0.59	New Zealand	4.867	27.494	0.549
	(0.134)	(0.298)	(0.009)		(0.315)	(0.629)	(0.013)
Austria	6.529	32.62	0.338	Norway	5.226	21.783	0.568
	(0.179)	(1.09)	(0.02)		(0.309)	(0.657)	(0.016)
Bulgaria	7.606	38.626	0.806	Peru	5.171	31.035	0.408
	(0.559)	(1.153)	(0.014)		(0.331)	(1.031)	(0.019)
Canada	6.001	29.22	0.389	Poland	6.898	33.825	0.579
	(0.17)	(0.788)	(0.021)		(0.222)	(0.676)	(0.016)
Czech Republic	6.753	39.153	0.546	Portugal	7.263	37.461	0.643
	(0.204)	(0.869)	(0.017)		(0.446)	(0.792)	(0.015)
Finland	4.953	28.304	0.396	Romania	6.075	30.321	0.527
	(0.283)	(0.768)	(0.014)		(0.583)	(1.07)	(0.023)
Germany	3.759	14.621	0.471	Slovakia	5.5	33.528	0.515
	(0.3)	(0.464)	(0.015)		(0.454)	(1.279)	(0.03)
Greece	5.323	21.915	0.656	Slovenia	6.352	37.269	0.574
	(0.355)	(0.775)	(0.023)		(0.497)	(1.867)	(0.047)
Iceland	5.485	30.458	0.533	South Africa	4.684	22.418	0.526
	(0.354)	(0.802)	(0.016)		(0.602)	(0.704)	(0.019)
Ireland	5.274	27.49	0.466	Sweden	6.326	20.73	0.498
	(0.445)	(1.046)	(0.023)		(0.155)	(0.833)	(0.018)
Israel	4.766	27.121	0.339	Switzerland	6.154	34.708	0.478
	(0.378)	(1.025)	(0.027)		(0.229)	(0.619)	(0.015)
Japan	4.183	17.48	0.34	United Kingdom	5.65	28.819	0.478
	(0.195)	(0.527)	(0.018)		(0.263)	(0.579)	(0.015)
Kenya	7.594	51.705	0.559	United States	7.065	17.651	0.579
	(0.367)	(0.904)	(0.018)		(0.113)	(0.395)	(0.013)

Table S5: Affective Polarization Estimates and Associated Standard Errors. Corresponds to main text Figure 8. Standard errors are in parentheses

S6.2 Full Correlation Results

Table S6 displays estimated correlations between affective polarization estimates and the additional variables discussed in the main text, along with the standard errors of those correlations. The estimates in Table S6 correspond to Figure 9 in the main text.

	Measure		
	Difference	Variance	CPC
Who in Power Makes Difference	0.043	0.013	0.022
	(0.028)	(0.013)	(0.022)
Vote Makes Difference	0.018	-0.208	0.402
	(0.022)	(0.014)	(0.022)
Very Close to Party	-0.175	-0.364	0.126
	(0.039)	(0.025)	(0.018)
Ideological Extremity	0.491	0.598	0.399
	(0.026)	(0.016)	(0.037)

Table S6: Correlates of Affective Polarization. Corresponds to main text Figure 9. Standard errors are in parentheses

S6.3 Weight of Features on CPC Estimates

Some of the simulation results presented in the main text and in section S3 suggest that the CPC may be more heavily influenced by one feature than the other, depending on the relative levels of each. To evaluate the relative weight of each feature on CPC estimates, I randomly vary each feature across a range of values that might be seen in real-world data, using the party affect data from above to define this plausible range of values.

To do so, I take data from three countries—Montenegro (lowest estimated polarization), Bulgaria (highest estimated polarization), and Ireland (approximate midpoint between Montenegro's and Bulgaria's polarization levels)—and calculate BSS and WSS separately for each. Using each estimate of BSS and WSS (low, medium, and high), I then take 1,000 independent draws from U($\frac{BSS}{2}$, $\frac{3BSS}{2}$) and U($\frac{BWSS}{2}$, $\frac{3WSS}{2}$). This exercise therefore allows me to randomly vary each feature *simultaneously* over a range of plausible values (something I could not do in the simulations above), while still ensuring that each feature varies independent of the other—an important attribute for determining the relative weight of each feature on the CPC, which is calculated from these two values.

I begin with descriptive evidence in Figure S10, which plots the value of each feature against the CPC value they combine to produce. As expected, the CPC increases as BSS increases and decreases as WSS

increases. The relationship between each feature and the CPC appears more linear than in some of the simulations above, but there does appear to be some evidence of diminishing returns to BSS at high levels of polarization and to WSS at low levels of polarization, consistent with simulation results.

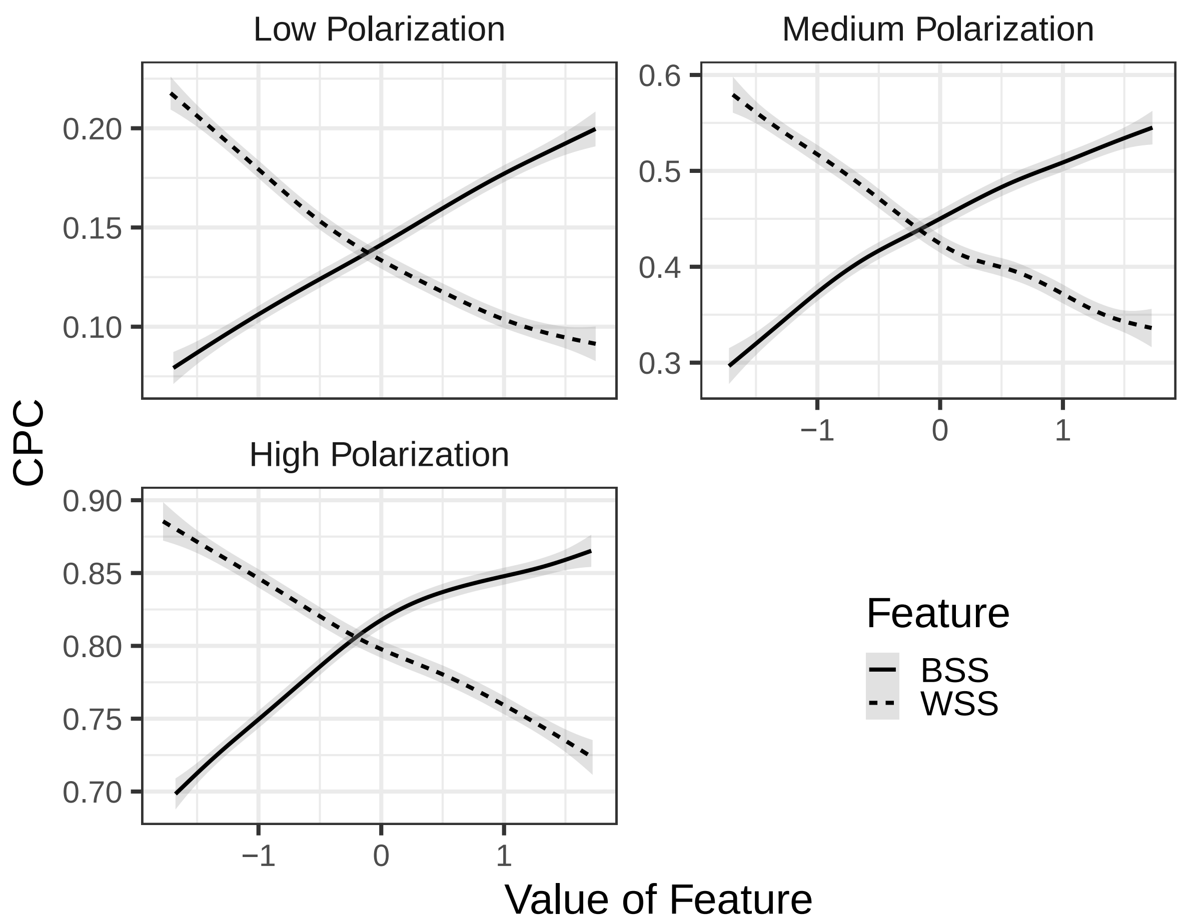


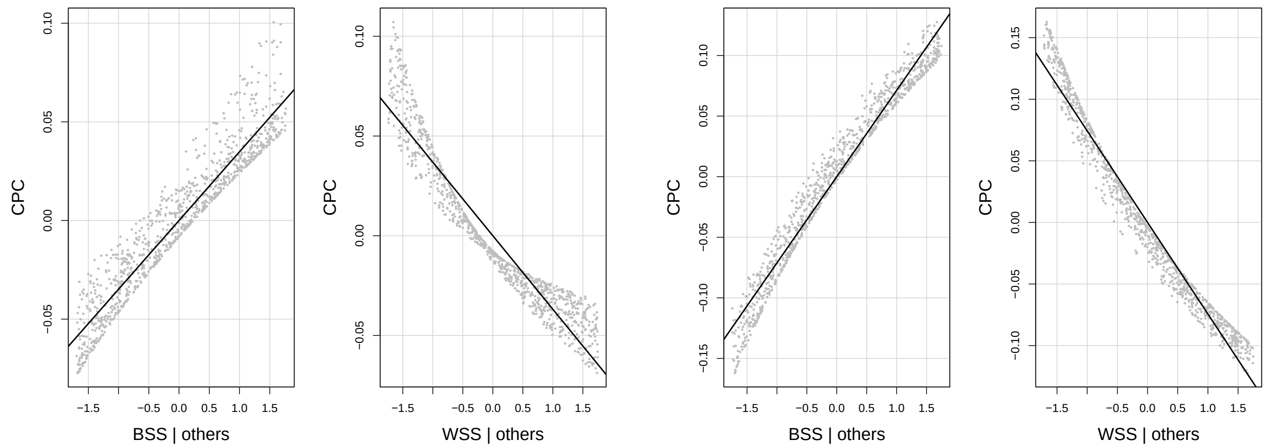
Figure S10: Changes in CPC Estimates with Independent Changes in Individual Features. Values of features unit-normalized to enable comparison.

More important, though, is to understand the *ceteris paribus* effect of each feature on the CPC. To investigate this, I fit OLS models of CPC estimates on BSS and WSS. Table S7 presents the results of these models. The coefficient magnitudes on BSS and WSS are very similar in all three models, giving some reassurance that the CPC is not being overwhelmingly driven by one feature more than the other. However, the faint diminishing returns observed in Figure S10 also show up in these results. The absolute magnitude of the WSS coefficient is greater than that of the BSS coeffiicient in cases of low and mid-range polarization. But when the overall level of polarization is high, that relationship flips, and the absolute magnitude of the BSS coefficient is greater than that of the WSS coefficient.

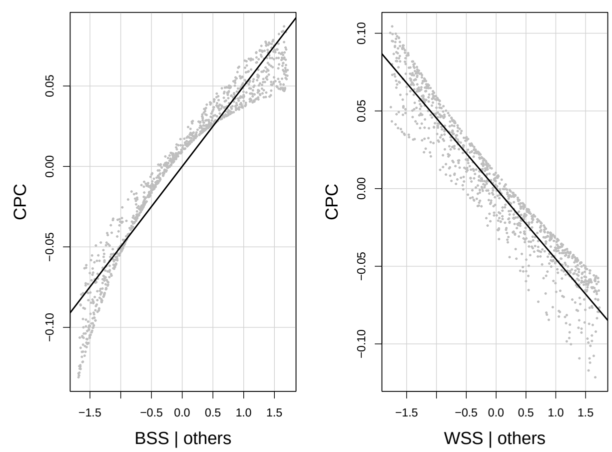
	Dependent variable:		
	CPC		
	Low Polarization	Medium Polarization	High Polarization
	(1)	(2)	(3)
BSS	0.035*	0.071*	0.050*
	(0.0004)	(0.0004)	(0.0004)
WSS	−0.037*	−0.074*	−0.045*
	(0.0004)	(0.0004)	(0.0004)
Constant	0.141*	0.441*	0.801*
	(0.0004)	(0.0004)	(0.0004)
Observations	1,000	1,000	1,000
R^2	0.950	0.981	0.959
Adjusted R^2	0.950	0.981	0.959
Residual Std. Error (df = 997)	0.012	0.014	0.014
F Statistic (df = 2; 997)	9,493.926*	26,295.530*	11,672.950*
Note:			*p<0.05

Table S7: Effect of Individual Features on CPC Estimates. BSS and WSS unit-normalized. Standard errors in parentheses.

These diminishing returns and their dependence on overall polarization can be seen more clearly in the added-variable plots in Figure S11. These plots use the models from Table S7 to assess the impact of each feature on the CPC while holding the other feature constant. Plots are separated into pairs according to low-, medium-, and high-polarization contexts. Within each pair of plots, the facet on the left displays the effect of BSS on the CPC, given WSS. The facet on the right displays the effect of WSS on the CPC, given BSS. Figure S11, plot (b) shows an approximately linear effect for both BSS and WSS on the CPC at a medium level of polarization. Just to the left, in Figure S11, plot (a), this linear effect appears again in the case of BSS, but WSS displays a diminishing effect, suggesting that BSS carries greater weight in the calculation of the CPC when the overall level of polarization is low. The opposite is true in Figure S11, plot (c), which reveals a linear effect of WSS on the CPC, but a diminishing effect of BSS, suggesting that the former carries greater weight when the overall level of polarization is high. However, in neither case does this diminishing effect appear to be substantial.



(a) Low Polarization (b) Medium Polarization



(c) High Polarization

Figure S11: Added-Variable Plots of Polarization Features and CPC Estimates. Values of features unit-normalized to enable comparison.

Taking into account all results presented in this section and in the simulations, the preponderance of evidence points toward three conclusions: First, each feature contributes approximately equally to CPC estimates, at least over a range of values observed in real-world data. Second, BSS may contribute more heavily in cases of low polarization, but this difference is not substantial. Third, WSS may contribute more heavily in cases of high polarization, but this difference is, again, not substantial.

References

Bradley, Ralph Allan and Milton E. Terry (Dec. 1952). "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons". In: *Biometrika* 39.3/4, p. 324.

Magee, Lonnie (Aug. 1990). "R2 Measures Based on Wald and Likelihood Ratio Joint Significance Tests". In: *The American Statistician* 44.3, pp. 250–253.