Published in final edited form as:
Annu Rev Sociol. 2022 July ; 48(1): 43–63. doi:10.1146/annurev-soc-030420-015435.

Measuring Ethnic Diversity

Liza G. Steele[1], **Amie Bostic**[2], **Scott M. Lynch**[3], **Lamis Abdelaaty**[4]

[1]Department of Sociology, John Jay College of Criminal Justice, and The Graduate Center, City University of New York, New York, NY, USA

[2]Department of Sociology, University of Texas, Rio Grande Valley, Brownsville, Texas, USA

[3]Department of Sociology, Duke University, Durham, North Carolina, USA

[4]Department of Political Science, Maxwell School of Citizenship and Public Affairs, Syracuse University, Syracuse, New York, USA

Abstract

Researchers have investigated the effects of ethnic heterogeneity on a range of socioeconomic and political outcomes. However, approaches to measuring ethnic diversity vary not only across fields of study but even within subfields. In this review, we systematically dissect the computational approaches of prominent measures of diversity, including polarization, and discuss where and how differences emerge in their relationships with outcomes of interest to sociologists (social capital and trust, economic growth and redistribution, conflict, and crime). There are substantial similarities across computations, which are often generalizations or specializations of one another. Differences in how racial and ethnic groupings are constructed and in level of geographic analysis explain many divergences in empirical findings. We conclude by summarizing the type of measurement technique preferred by outcome, when relevant, and provide considerations for future researchers contemplating how best to operationalize diversity. Finally, we highlight two less widely used yet promising measures of diversity.

Keywords

ethnic diversity; ethnic fractionalization; ethnic heterogeneity; race; ethnicity

1. INTRODUCTION

Ethnic diversity—heterogeneity of racial and ethnic groupings within a country—is a central concept in scores of academic studies (e.g., Alesina et al. 2003, Fearon 2003, Habyarimana et al. 2007, Van der Meer & Tolsma 2014) in addition to being an implicit or explicit component of high-stakes contemporary policy debates worldwide. However, the operationalization of ethnic diversity varies widely across studies. Moreover, varying approaches to measuring ethnic diversity can yield critical differences in explaining social,

LSteele@jjay.cuny.edu .

DISCLOSURE STATEMENT
The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

political, and economic outcomes in quantitative analyses (Desmet et al. 2009, Steele 2016, Steele & Abdelaaty 2019). Yet there is no recent source that provides a broad overview of how diversity is measured.

In this review, we fill that gap by examining existing quantitative approaches to measuring ethnic diversity using all prominent cross-nationally comparable measures. First, we compare the mathematical formulae used in the calculation of various indices and discuss how different measures are related to one another as special or general cases. Then, we examine how these measures are employed across empirical social science studies. We analyze the results of studies in which ethnic diversity is used to predict various outcomes of interest to sociologists (social capital, generalized trust, economic growth, public-goods provision, conflict, and crime) and delve into the ways in which differences in measurement of the concept may affect findings.

As we show below, measures of ethnic diversity are largely similar to one another mathematically. And indeed, they ought to be broadly similar to the extent that they are seeking to capture comparable concepts. However, as our review of the literature demonstrates, differences in underlying data sources, level of aggregation, and coding between the different measures are consequential in driving divergent results.

Within the literature on ethnic diversity, a substantial body of literature specifically focuses on the effects of immigration and immigrant-based diversity. While immigration is one important vector of diversity, this review is generally focused on broader measurements. However, we highlight a few prominent studies on immigration, especially those that examine how the effects of immigration operate differentially depending on existing diversity in immigrant-receiving societies. For thorough reviews of the effects of immigration, readers are directed to Portes & Vickstrom (2011) and Van der Meer & Tolsma (2014) in previous volumes of this journal.

2. KEY APPROACHES TO MEASUREMENT

Measures of diversity go by a variety of names in the literature, and the names vary depending on the specific characteristic(s) on which they are based, such as language, race, or ethnic groupings represented within a geographic area. Thus, measures referred to in the literature as ethnic fractionalization and linguistic fractionalization may involve the same fundamental calculations, but they are named differently simply because the computation is applied to distinct variables and data. Indeed, in the literature, many scholars claim to have developed new measures of diversity, but these new measures are often identical computationally to other, previous measures and have simply been applied to a new data source or variable (e.g., Alesina et al. 2003). An overview of differences in data sources is provided by Marquardt & Herrera's (2015) review.

Table 1 lists the most commonly used measures of diversity and polarization with common notation, in the order they are discussed below. Here, we present a nontechnical overview of the measures. The Supplemental Appendix discusses in detail the computations underlying each measure presented in the table and shows their similarities and differences.

There are two general classes of measures in the literature: measures of diversity (or fractionalization) and measures of polarization. Both types of measures aim to capture the extent to which members of a population differ from one another, but the measures are distinct in their emphases. Measures of fractionalization emphasize the number of groups within a population, while measures of polarization emphasize the compositional imbalance of groups within a population (Koopmans & Schaeffer 2015). Both types of measures may incorporate quantification of the extent of differences between groups (i.e., disparities), and they can be related (i.e., shown to be general or special cases of one another).

Both types of measures are macrolevel measures, because diversity is a characteristic of populations. They are incorporated as predictors, however, in analyses of both micro- and macrolevel outcomes in the literature, and they are used across a wide range of research areas, which has led to a wide diversity in nomenclature and notation despite many underlying similarities among the measures.

A basic way to measure diversity is simply to record the proportion of a population in defined ethnic, linguistic, or other groups. Most measures of immigrant populations do so by including a single measure of percent foreign born. However, when measuring diversity more broadly, there are two major problems with that approach. First, ethnic groups vary across countries, so using the same set of ethnicities across countries would result in many unobserved groups. This problem can easily be remedied by simply measuring the proportion of the population in the largest ethnic group in a country, whatever it is. One can also include the proportion in the smallest group, along with, perhaps, the number of ethnic/linguistic groups that exist in the population (Ellingsen 2000). However, the second major problem with measures of percentages in ethnic groups is that including multiple measures simultaneously makes for difficult interpretation regarding the effect of diversity, because both the number of groups and their relative sizes indicate something about heterogeneity. A single measure would arguably be better, but using a single measure also turns out to be problematic, as we discuss.

One of the simplest single measures of diversity is the product of the proportions of the population in each group in the population. Whereas the proportion in the largest group represents increasing heterogeneity only up to a point—beyond which a large proportion indicates reduced heterogeneity—the product of the proportions of persons in each group is maximized when each group in a population is of equal size. For example, for a population that consists of two groups, this product-of-proportions measure reaches a maximum when both groups comprise 50% of the population. In a population with three groups, this measure is maximized when each group comprises one-third of the population, and so on. We might call such a measure the product-of-proportions (PP).

PP, while understandable, has no inherent theoretical interpretation, because the specific maximum varies depending on the number of groups involved in the calculation, which may differ across populations within a sample. An alternative measure with a theoretical basis—and which is most common in the literature—is the complement of the Herfindahl index, also called the Herfindahl-Hirschman index (HHI). The HHI was developed as a measure of industry concentration and is computed as the sum of the squared shares of each group in the population. Larger values indicate that the market is controlled by a limited number of firms. For example, if one firm has 100% of a market, HHI = 1, indicating a monopoly. If, in contrast, there are 100 firms in a market, and each has a 1% share, then HHI = $100(0.01)^2$ = 0.01, indicating a very low concentration.

When shares of a population reflect ethnic or linguistic group percentages, the complement of the HHI (1 – HHI) is a measure of diversity, rather than concentration, and often is called ethno-linguistic fractionalization (ELF) in the literature reviewed here, because the grouping characteristics are usually defined by races/ethnicities and/or languages. The interpretation is that ELF represents the probability that two people selected at random from a population belong to different groups. When there are only two groups, ELF is proportional to (i.e., twice) the primitive PP measure discussed above. However, with three or more groups, ELF's interpretation is clearer than the PP measure.

Although ELF has a strong theoretical underpinning and is the most commonly used measure of diversity in the literature, it has some well-known limitations. A major shortcoming is that it assumes that differences between groups in a population are equivalent: ELF makes no distinction between two groups that are somewhat similar linguistically, ethnically, or otherwise, versus two groups that are substantially different. For example, in the United States, non-Hispanic whites, black Hispanics, and black non-Hispanics would all contribute equally to the population's overall heterogeneity based on ELF, even though black Hispanics and black non-Hispanics likely share more in common than either group shares with non-Hispanic whites.

A second shortcoming is that ELF is unidimensional in terms of how groupings are commonly measured. For example, although the measure is called ethno-linguistic fractionalization, in fact, groupings are generally quantified on a single characteristic, such as race. Collier & Hoeffler (2004) provide an exception. They created separate measures of ethnic and religious fractionalization and constructed a single measure from them by multiplying them together and then adding the larger of the two to this product.[1] Okediji (2005) provides another exception: He used a three-way computation of ELF involving race, religion, and language subgroups.

A third significant shortcoming of ELF is that it cannot capture the extent of heterogeneity (or homogeneity) within groups nor the extent of differences between them. Moreover, ELF cannot distinguish between populations that are characterized by having a large politically dominant group and a small politically weak group versus a large but politically weak group and a small politically dominant group. For example, as Abascal & Baldassarri (2015) show, an area in the United States that is 80% white and 20% black has the same ELF as one that is 20% white and 80% black, although, as the authors point out, such areas would surely be different in many ways.

Arguably, a good measure of diversity should take into account not only the number of groups in a population and their relative share of the overall population, but also

[1]The rationale for adding in the larger of the two was that, if either measure were 0, the product is 0, thus understating heterogeneity.

how different groups are from one another along more than one dimension, and perhaps capture some characteristic of power differentials between groups. Several generalizations or extensions of ELF that address the limitations discussed above have been proposed and are commonly seen in the literature. These include Baldwin & Huber's (2010) cultural fractionalization (CF) index and Bossert et al.'s (2011) generalized ethno-linguistic fractionalization (GELF). As described below, the extension of ELF to CF and then to GELF at least partially resolves the first two limitations of ELF listed above but does not resolve the third.

A first generalization of ELF involves the incorporation of a similarity measure, s, into the calculation (see Table 1). Most recently, Baldwin & Huber (2010) presented an index of CF. This measure was first proposed by Greenberg (1956) and later used by Fearon (2003), who developed a measure of s using linguistic trees (see Baldwin & Huber 2010). The similarity measure, s, is assumed to take a value between 0 and 1, with 0 representing entirely dissimilar languages and 1 representing entirely similar ones. It is straightforward to see that CF reduces to ELF when s is dichotomous and takes a value of 0 when computing the contribution of pairs of different groups (see the Supplemental Appendix).

Bossert et al. (2011) provide GELF, an alternative extension to ELF. In contrast to ELF and CF, GELF is defined in terms of individual-level similarity matrices, S, where each element represents the overall similarity of each individual to other individuals. An important contribution of Bossert et al. (2011) is that they emphasize that the elements of the similarity matrix S are not constrained to represent a single dimension of similarity, such as language or ethnicity.

ELF and CF were specifically derived with ethnicity or language in mind as a lone measure of similarity. Multidimensional measures of group similarity are more difficult to derive because as the number of dimensions increases, group sizes become smaller. Thus, it may be difficult to derive stable group-based measures of similarity based on numerous characteristics. However, it is fairly straightforward to derive individual-level measures of similarity using some sort of multivariate distance measure, such as a Mahalanobis distance (although we find no examples of this in the reviewed literature). For example, one could compute the Mahalanobis distance for age, education, and income, and rescale the distance measure to the [0,1] interval to obtain a measure reflecting similarities of individuals on three characteristics. Thus, the individualized approach illustrated by GELF can potentially represent a more nuanced degree of heterogeneity than the group-based computations involved in ELF and CF.

Esteban & Ray (1994) begin to address the third limitation of ELF by arguing that extant measures of diversity do not tell us much about polarization, which they define as a tendency toward increasing homogeneity within groups in the face of stable or growing heterogeneity between groups. Indeed, inequality—as measured by fractionalization measures—may decrease at the same time that polarization is increasing.

They propose a measure that incorporates a parameter reflecting effective antagonism between groups, where the antagonism is defined both by affinity for members of one's in-group and distances between groups (the Esteban-Ray index, or ER).[2] If effective antagonism is fixed at 1, and the similarity measures are dichotomous, ER can be shown to be proportional to HHI (the complement of ELF; see the Supplemental Appendix for details). Following Esteban and Ray's development of a generalized polarization measure, others have developed similar measures that have somewhat unique features.

Reynal-Querol (2002) developed their own polarization index, and Desmet et al. (2005) developed a peripheral polarization measure. First, Reynal-Querol (2002) developed a measure (RQ) that centers polarization over a theoretical maximum value of one-half; the assumption is that the most polarized population is one with two equally sized groups. Thus, RQ polarization involves computing the deviation of each subpopulation from one-half and taking that value as a proportion of the maximum of one-half, as shown in Table 1. However, it does not consider distances between groups; it is simply based on an assumed in-group affinity of 1, meaning that group members identify strongly with one another. Desmet et al. (2005) developed a peripheral polarization measure that assumes the centrality of a majority subpopulation (group 0) and involves only distances between each minority group and the majority group.

Recently, Koopmans & Schaeffer (2015) developed a measure they call generalized entropy (GE) that can be reduced to measure either fractionalization or polarization. They argue that measures of diversity and polarization both attempt to represent three features of population heterogeneity: the extent of variety of population subgroups (i.e., how many subgroups exist), the extent of balance between the groups (how equally distributed are the groups), and the disparities between the groups (between-group heterogeneity, distances). They argue that both types of measures try to capture all three with a single number, but that diversity measures emphasize variety, while polarization measures emphasize balance. Thus, it may not be surprising that Montalvo & Reynal-Querol (2005) have argued that diversity measures may be better than polarization measures for predicting general inequality outcomes that rely on individual-level measures, such as the Gini coefficient for income inequality. (Indeed, the Gini coefficient is a special case of CF when the similarity/distance measure is simply the absolute value of the difference in incomes between two individuals.) In contrast, given polarization measures' emphases on balance, Montalvo & Reynal-Querol (2005) suggest that these are better than diversity measures for predicting civil conflict.

Koopmans & Schaeffer (2015) argue that GE can be used to compare the relative importance of variety and balance for a given outcome. As shown in Table 1, the measure has a parameter that captures polarization, but this parameter is also included in the denominator of an exponent within the formula, enhancing the interpretation of the GE. If the parameter is zero, the measure is simply a count of the number of groups in the population, emphasizing the extent of variation within a population but ignoring their relative balance (a pure diversity measure). If the parameter is equal to two, GE is simply the inverse of

[2]Effective antagonism encompasses the concepts of identification and alienation. Identification is the extent to which an individual identifies with other individuals in the same group given the size of the group, which may also be equal to 0 to capture the possibility of identification being independent of the size of the group. Alienation—how an individual in group A feels about an individual in group B—increases with the distance between groups and may be affected by that individual's identification with coethnics in group A (Desmet et al. 2009, Esteban & Ray 1994, Steele 2016).

the HHI. Finally, as the parameter increases toward infinity, it approaches the inverse of the proportion of the population in the largest group, reflecting the extent of balance in a population.

Also in the past few years, an additional concern has been raised regarding the measurement of heterogeneity. Above, we noted that the unidimensionality of most diversity measures is a limitation, in terms of understanding their effect on macrolevel outcomes. Some scholars focused on predicting conflict (e.g., civil wars) have tended to prefer polarization measures, while the literature focused on predicting inequality of various outcomes (e.g., income inequality) has tended to prefer fractionalization measures. The rationale for the preference of polarization measures in the former literature is that in-group affinity as well as between-group differences matter for producing conflict. A neglected concern in that literature, as noted by Lieberman & Singh (2012), is that of the cross-cuttingness of cleavages between groups. For example, consider two populations in which there are two fundamental dimensions of difference—say, language and ethnicity. In each society, half speak language a, and half speak language b, and each half includes members of both ethnic groups c and d. ELF measures constructed on the basis of either characteristic (language or ethnicity) will be equal for both societies, but suppose that in one society, 25% are in each of the four potential cells implied by the two dimensions (ab, ac, bc, bd), while in the other, the two characteristics align so that 50% are in cell ac and 50% are in cell bd. Lieberman & Singh (2012) argue that the latter society would be more prone to conflict, because the cleavages are not cross-cut. Instead, the different dimensions along which the society is divided align, augmenting polarization. However, it is not clear that the polarization measures shown above can capture this type of difference between societies, even if a multidimensional distance measure were used.

The most commonly used measures of diversity and polarization tend to be highly correlated, and even when they are not, they often can be made to be so with simple transformations (e.g., Marquardt & Herrera 2015, Montalvo & Reynal-Querol 2005). Thus, their differential influences on any outcome are often simply a matter of functional form. For more detailed summaries of measures than we can provide here, we recommend the articles by Bossert et al. (2011), Desmet et al. (2009), Esteban & Ray (2011), Lieberman & Singh (2012), and Reynal-Querol (2002).

Despite mathematical similarity between measures, the diverse data sources to which they are applied explain many of the disparities in findings in the empirical literature. A final concern is that results may also vary across studies that use different levels of aggregation because of the modifiable areal unit problem (MAUP).[3] Although we do not compare results of studies with respect to geographic unit size or shape here, it is possible that differences in geographic delineations may help explain variations in findings, over and above the choice of data and measures.

[3]MAUP is a well-known issue in macrolevel analyses. In brief, aggregate measures are constructed by grouping individuals together and computing a desired quantity summarizing the group, such as the proportion over age 50. The values of these quantities, and the relationships between quantities, may very well vary if the size of groupings is altered or if the geographic boundaries of the grouping are adjusted. Changing the unit of analysis from counties to states is an example of the former kind of change in grouping, while gerrymandering congressional districts is an example of the latter (see Wong 2009 for further discussion).

3. APPLICATIONS IN THE SCHOLARLY LITERATURE

In this section, we summarize major studies employing the measures introduced above. We review literatures on social capital and generalized trust, welfare states and economic growth, conflict, and crime. Most studies of social capital and trust have operationalized diversity as ELF (1 – HHI), although there are debates about which level of measurement is most salient; a subset of this literature focuses on immigration, almost always measured simply as percent foreign born in a population.

Similarly, within the welfare state and economic growth literature, most studies have measured diversity using ELF, though some have also considered polarization (e.g., Desmet et al. 2009). Overall, however, the major point of debate regarding measurement centers around the best way to categorize ethnic groupings and how and whether to take distances between groups into consideration. Again, a subset of this literature focuses on immigration using percent foreign born as the measure.

Measures of polarization have also been linked to conflict onset and incidence, while heterogeneity appears to have no effect. However, some conflict scholars emphasize the importance of focusing on politically relevant cleavages in particular, generally encapsulated through unique data or coding rather than unique measures.

Finally, ethnic diversity has long been hypothesized to increase crime; however, the empirical results have been mixed. While ELF (1 – HHI) has been the dominant measurement of ethnic heterogeneity in this literature, the appropriate level of aggregation has been a source of contention, and results tend to vary depending on which level of aggregation is chosen.

Below, to simplify cross-referencing with Table 1, we refer back to the measures therein using their line numbers. We focus on studies that are cross-national, except where single-country studies (usually studies of the United States) have been used as the basis for broad generalizations elsewhere in the literature.

3.1. Social Capital and Generalized Trust

The past 15 years have seen a great deal of interest in the relationship between ethnic diversity and social capital and/or generalized trust. In fact, this subject has already been the focus of two reviews in this journal (Portes & Vickstrom 2011, Van der Meer & Tolsma 2014) and another two in the year 2020 alone (Baldassarri & Abascal 2020, Dinesen et al. 2020). Thus, we discuss here a few of the key studies from this vast and already well-reviewed literature to illustrate how the different measures discussed above have been used.

Much of the debate around this topic accelerated after the publication of Putnam’s (2007) seminal article, in which he advances the sweeping claim that people in more diverse communities trust one another less. Putnam analyzes data from the 2000 Social Capital Community Benchmark Survey of US residents. He models only one outcome, trust in neighbors, and finds it to be inversely related to heterogeneity; in fact, the measure he uses is HHI (Table 1, measure 4), which measures homogeneity, but discusses the results in terms of implications for heterogeneity. Among the many empirical examinations inspired by Putnam's theories, a scholarly consensus emerges that he neglected to consider structural factors that are more relevant to the development of social capital than ethnic diversity.

In an analysis that is more global in scope, Kesler & Bloemraad (2010) consider the effects of immigration on social capital in 17 to 19 advanced democracies.[4] Data on immigration, measured as percent foreign born of the total population, come from the United Nations (2005), while the social capital measures (trust, civic engagement, and political participation) come from the World Values Survey (Eur. Values Study Group & World Values Surv. Assoc. 2006).[5] They find that the effects of immigration on social capital are not consistent across measures, and both low levels of income inequality and strong multicultural policies mitigate any negative effects of immigration.

Building on this emphasis on inequality, social capital, and ethnic heterogeneity findings, Portes & Vickstrom's (2011) review further highlights the importance of inequality in determining the salience of ethnic diversity. Because theirs is a review article, various approaches to measuring diversity are represented, but most studies the authors review use immigration as percent foreign born. They point to the consensus that had emerged in the empirical literature that when intergroup contact is high and economic inequality is low, indicators of civic engagement or trust do not decline. However, in contexts of high inequality and spatial segregation, empirical findings suggest that ethnic diversity, operationalized using various measures across the literature they review, is associated with lower social capital. Their ultimate conclusion is also that structural factors are preeminent, with social capital emerging as a by-product of such circumstances.

The empirical findings from Abascal & Baldassarri (2015) are consistent with the conclusions of Portes & Vickstrom (2011) and Kesler & Bloemraad (2010). In a thorough and direct reexamination of Putnam's (2007) analysis, Abascal & Baldassarri (2015) use the same Social Capital Community Benchmark Survey data set to demonstrate that ethnoracial, residential, and economic differences between communities and the residents who select into living in them are much better explanations of individual variation in self-reported trust and cooperation than is ethnic diversity. They examine five different measures of trust —generalized trust, trust in neighbors (used by Putnam), in-group trust, out-group trust, and perceptions of neighborhood cooperation, and find that ELF, which they call HHI_{hetero}[6] (measure 3), is only inversely related to trust in neighbors. They conclude that "diversity is a negligible predictor of trust compared with classic sociological indicators of inequality" (Abascal & Baldassarri 2015, p. 752).

Van der Meer & Tolsma (2014) also conclude in their large-scale review that spatial segregation, highlighted by the Portes & Vickstrom (2011) review, is a critical aspect of the

[4]They analyzed residents of advanced democracies for which they had multiple waves of data, which varied between 17 and 19 countries, depending on the outcome variable.

[5]Social capital was operationalized via the following measures: (*a*) generalized trust, (*b*) six types of organizational membership, and (*c*) nonelectoral political actions (signing a petition, joining a boycott, and attending a lawful demonstration).

[6]They use the measure 1 – HHI to represent the concept of heterogeneity as opposed to homogeneity.

diversity-social capital relationship. In fact, they describe the combination of heterogeneity and inequality, particularly in conjunction with segregation, as a "potentially explosive mix" (Van der Meer & Tolsma 2014, p. 474). Yet, overall, they conclude that there is no fundamental association between racial/ethnic heterogeneity and social cohesion. The exception may be the neighborhood level, particularly in the context of the United States. Again, because this is a review article, a range of measures are included in their analysis—also true for the next two studies we discuss.

Also highlighting the importance of the neighborhood level of analysis, Dinesen et al.'s (2020) review is unique among the four in concluding that there is a more general negative relationship between ethnic diversity and social trust, although they caution that sizes of the effects are, at most, moderate, and that the alarm about diversity eroding cohesion is, thus, overstated.[7] Their analysis demonstrates that among the four categories of trust they examine, the negative relationship is most robust between trust in neighbors and diversity at the neighborhood level. Notably, they find that the salience of diversity decreases as the level of measurement broadens from neighborhood to region/municipality to country, with it being nonsignificant at the country level. Dinesen et al. (2020) conclude that the relationship between ethnic diversity and social trust is only slightly attenuated, and remains negative and statistically significant, when controlling for the potential confounders and mediators that other reviews had shown to explain away the effects.

However, as Baldassarri & Abascal (2020) emphasize in their review, homogeneous and heterogeneous communities in Western societies differ in systematic ways that limit the validity of attributing observed variations between them to ethnic diversity itself. Research on Asian and African countries, which have some of the greatest diversity among their native populations, is extremely limited. Thus, we should be more conservative about faulting ethnic diversity for divisions that reflect new and rapid changes in a specific part of the world. Another critical point raised by Baldassarri & Abascal (2020) is the importance of contact in diverse workplaces, which has been less widely studied in sociology than diversity in countries, regions, and neighborhoods.

Additional challenges to understanding this relationship pertain to the cross-sectional nature of most empirical studies on diversity and social capital and questioning whether existing measures of social capital best encapsulate social cohesion. Ramos et al.'s (2019) longitudinal examination of the effects of religious heterogeneity on perceived quality of life raises questions about the validity of drawing conclusions about the effects of diversity on social outcomes using a snapshot at a single point in time. They measure religious diversity, effectively ELF (1 – HHI, measure 3), and extract data on perceived quality of life from the World Values Survey, European Social Survey, and Latinobarometer to analyze 22 years of worldwide data. While they find negative reactions to increased diversity (lower self-reported quality of life explained by decreased trust in others) in the short term, the long-term benefits of intergroup contact ultimately alleviate these negative influences.

[7]They conduct a meta-analysis of 1,001 estimates from 87 studies (encompassing wide variation in approaches to measurement of and data sets on diversity) and divide trust into four different categories: generalized social trust for strangers, out-group trust for members of salient ethnic out-groups, in-group trust for members of the same ethnic group (fellow natives), and trust in neighbors for those in the same residential environment.

While the results of these studies are inconsistent, much of this is driven by the level of aggregation, especially whether the authors are looking at local heterogeneity or national heterogeneity. Yet, across all of these studies, there emerges a consensus that trust in neighbors has a unique relationship to diversity, and that diversity is most salient when measured at the local level. However, we agree with those who caution against using these findings to make broad claims given the small effect sizes, consistent failure to consider critically important contextual factors (particularly socioeconomic and demographic variables), the lack of longitudinal data on diversity, and the excessive focus on a small number of countries that vary relatively little from one another in terms of diversity (e.g., Abascal & Baldassarri 2015, Van der Meer & Tolsma 2014). When the foreign-born population is the not the focus of social capital studies, ELF (measure 3) is the measure most likely to be used. Because ELF is so widely used, it is logical that so many social capital and trust researchers select it for their analyses. However, the results of such studies should be interpreted through the lens of the limitations outlined in Section 2.

3.2. Economic Growth, Inequality, and Redistribution

In contrast to the literature on social capital, during the 1990s, the development literature was dominated by questions about Africa. Specifically, many wondered why Africa had failed to grow economically, stabilize politically, and build suitable infrastructure, while East Asia had generally been successful. Most pointed to policy failures; however, the cause of these policy failures remained unclear. Some researchers suggested that it was the product of diversity in a society.

In their seminal article, Easterly & Levine (1997) claim that the source of these policy failures and lack of growth is ethnic heterogeneity. Using a measure of ELF (measure 3) from the *Atlas Narodov Mira* (Bruk et al. 1964), they find strong evidence supporting the hypothesis that ethnic heterogeneity decreases growth directly and increases factors associated with slow growth. Even when they test their hypothesis using four other measures of ethnic heterogeneity, their findings are consistent. In fact, they find that "ethnic diversity alone accounts for about 28% of the growth differential between the countries of Africa and East Asia" (Easterly & Levine 1997, p. 1207). Based on the findings of this article, discussions of ethnic heterogeneity emerged as central to cross-national research on economic growth and welfare state development. However, the measurement of ethnic heterogeneity has been hotly debated in the literature and is often credited as the basis for inconsistent findings.

In a critique of the Easterly & Levine (1997) article and the literature it engendered, Alesina et al. (2003) question the ethno-linguistic groupings used to calculate heterogeneity. Using the same formula (ELF, measure 3), they calculate heterogeneity from alternative data sources, drawing on separate data to calculate indices of ethnic, linguistic, and religious heterogeneity. Yet with these new data, they largely confirm the findings of Easterly & Levine (1997): Ethnic and linguistic heterogeneity decrease growth and many of the policies associated with it. While the findings of this article make it noteworthy, the major contribution of the article is the data on heterogeneity calculated and provided by the authors, including information on underlying groupings. Since then, many researchers have relied on these data to conduct their analyses (e.g., Churchill & Laryea 2019).

Stichnoth & Van der Straeten (2013) provide a detailed review of the literature spawned by the Alesina et al. (2003) paper and data set. Specifically, they review the literature examining the effects of ethnic diversity on actual public spending as well as preferences for this kind of spending. They find that most cross-national studies identify a negative relationship between ethnic heterogeneity and social spending generosity. However, at the subnational level, the results are more ambiguous. For example, while overall spending may not be affected, spending on welfare or public goods is negatively associated with ethnic heterogeneity.

Most scholars argue that ethnic heterogeneity negatively affects social spending through individual preferences. More heterogeneous societies may be more prone to rent-seeking and less likely to reach consensus over public spending. Yet most studies are observational and utilize survey data. More recently, however, experimental studies have attempted to isolate the mechanism driving the relationship between ethnic diversity and preferences for social spending by utilizing trust games. For example, Habyarimana et al. (2007) use data from experimental games conducted in a highly diverse Ugandan slum to see if participants exhibit preferential treatment toward other players from the same ethnic background. They consider three possible explanations for why one might enact coethnic preferential treatment: personal in-group preferences, networks/social capital, and consideration of sanctions from coethnics. Interestingly, their results suggest that it is not individual discrimination or ethnic social capital that influences preferences, but rather fear of sanctions from other members of their ethnic group.

In contrast, Desmet et al. (2009) critique predominant measures due to their omission of distance between groups. They argue that the reason why the effects of linguistic diversity on redistribution are so often not statistically significant is that they fail to account for distance. In fact, they find that ELF, ER, and RQ (measures 3,7, and 8), and their own peripheral diversity measure (measure 9), all reach statistical significance when distances between groups are included in their calculations, indicating a negative effect on redistribution consistent with previous literature.

In a final influential study, Baldwin & Huber (2010) contend that while the literature indicates that ethnic diversity makes governance more difficult, these results should be questioned due to the omission of between-group inequality (BGI, a group-based computation of the Gini coefficient; measure 2). The authors compare the effects of BGI with ELF and CF (measures 3 and 5) on public-goods provision. Contrary to previous literature, they find no robust relationship between ELF or CF and public-goods provision, yet find a robust and sizable negative relationship between this public-goods provision and their measure of BGI.

Within the literature on ethnic diversity and preferences for redistribution, a substantial body of literature specifically focuses on the effects of immigration and immigrant-based diversity. While this literature almost always measures immigration as percent of the population that is foreign born, the findings of this research have been mixed. Some have suggested that the size of the immigrant population increases support for redistribution (Brady & Finnigan 2014, Finseraas 2012), while others have found that it decreases support (Mau & Burkhardt 2009, Schmidt-Catran & Spies 2016), and some have found no relationship at all (Hainmueller & Hiscox 2010, Senik et al. 2009, Steele 2016). Many have offered more qualified assessments, claiming that immigrant qualities (Reeskens & Van Oorschot 2012) or economic conditions (Eger & Breznau 2017) influence the relationship, or that regions, not countries, are the appropriate level of analysis. For example, both Eger & Breznau (2017) and Alesina et al. (2021) find that immigration and immigrant residential segregation, respectively, decrease support for redistribution at the subnational level. Eger & Breznau (2017) further critique the extant literature for its generally limited examination of only a small number of Western European countries, and Breznau et al. (2021), through a crowdsourced replication study, reveal just how sensitive results are to the ways in which immigration is measured.[8] Finally, some research suggests that it is not actual immigration that influences redistribution attitudes, but how immigration is perceived and experienced (e.g., Semyonov et al. 2008, Steele & Perkins 2019), and that opposition to immigrants' social rights may be more an indicator of antiforeigner sentiment than a reflection of welfare state preferences (Eger & Breznau 2017, Scheepers et al. 2002).

For the past 25 years, the literature on development and redistribution has investigated the effects of general ethnic heterogeneity and the size of the immigrant population on economic growth, the development of the welfare state, and attitudes about the welfare state. Generally, the findings have indicated that ethnic heterogeneity hinders economic growth and the development of generous welfare-state institutions at the local, regional, and national levels. However, despite fairly consistent findings, controversy continues to proliferate in this literature about how best to measure ethnic heterogeneity. Nonetheless, most studies use a version of ELF but vary how categories are constructed or which data source is employed. ELF is the optimal measure to enhance comparability with previous findings; however, if power dynamics or similarity between groups are of theoretical importance incorporating BGI is recommended.

3.3. Conflict

One of the largest bodies of scholarly work on the effects of ethnic diversity pertains to conflict. The empirical literature has diverged on whether ethnic divisions lead to intrastate conflict, depending on the measure of ethnic diversity used. However, because the focus is on civil war, these measures are almost always limited to the national level.

On the one hand, ethnic heterogeneity does not appear to be robustly associated with civil war onset. Using ELF (measure 3) and an analogous measure of religious fractionalization, Fearon & Laitin (2003) show that diversity does not have a statistically significant effect on the probability of civil war onset in the post–Cold War period, arguing that the risk for civil war is tied to conditions that favor insurgency (like poverty, instability, rough terrain, and large population). Collier & Hoeffler (2004) argue that diversity may reduce the viability

[8]Types of measurement include stock, flow, change in flow (the derivative of flow), using various fractionalization indices, or some combination thereof.

of rebellion by limiting rebel recruitment to a single group; they demonstrate that social fractionalization, which they code based on ELF (measure 3, relying on *Atlas Narodov Mira* data) and a similarly computed religious fractionalization index using data from Barrett (1982), decreases the probability of rebellion so long as societies avoid ethnic dominance (see also Elbadawi & Sambanis 2000).[9]

In contrast, Annett (2001) finds a correlation between his original measure of social fractionalization (based on measure 3's formula)[10] and a wide-ranging proxy for political instability. Vanhanen (1999) also finds a positive association between the ethnic heterogeneity index he constructs[11] and ethnic violence, albeit using only cross-tabulations and univariate regressions. In a sensitivity analysis that includes ethnic fragmentation indices drawn from the work of Fearon & Laitin (2003), Collier & Hoeffler (2004), and Vanhanen (1999), Hegre & Sambanis (2006) conclude that ethnic fractionalization is robustly correlated with the onset of low-intensity violence, but not full-scale civil war.

On the other hand, ethnic polarization seems to be an important predictor of conflict onset. Desmet et al. (2012) create a unique measure of diversity using language trees. They find that linguistic polarization (calculated according to measure 8) and fractionalization (calculated according to measure 3) are associated with the onset of civil conflict, though only at the highest level of aggregation of linguistic cleavages. Østby (2008) argues that the probability of civil conflict onset rises when ethnic polarization coincides with social disparities (horizontal social inequality). After coding their own dichotomous ethnic polarization measure, Forsberg (2008) finds that a dichotomous ethnic polarization measure shapes contagious processes for ethnic conflict: When one state experiences conflict, neighboring states that are polarized are more likely to see an outbreak of conflict. Collier & Hoeffler (2004) show that ethnic dominance (i.e., whether a single ethnic group makes up 45–90% of the total population) is associated with civil war onset. Hegre & Sambanis (2006) corroborate this finding, while also demonstrating that dominance does not affect the risk of lower-level conflict.

Ethnic polarization has been linked to civil war incidence as well. Using the Reynal-Querol (2002) polarization index (measure 8), Montalvo & Reynal-Querol (2005) find a positive effect of ethnic polarization on the incidence of civil war. Esteban et al. (2012), relying on a measure of polarization based on Duclos et al.'s (2004) measure 7, also find that ethnic polarization is positively associated with the incidence of conflict. Ellingsen (2000) measures multiethnicity in several ways and finds that the incidence of domestic armed conflict is higher when the largest ethnic, religious, or linguistic group represents less than 80% of the total population. Meanwhile, the number of groups and the size of the largest minority group relate to the incidence of conflict in the shape of an inverted-U curve. Relying on Gurr's (2009) Minorities at Risk data, Toft (2005) emphasizes settlement patterns and concludes that rebellion is more likely when an ethnic group is concentrated

[9]However, in a later paper based on different data, Collier et al. (2008) find that social fractionalization increases the risk of civil war.
[10]Annett (2001) calculated social fractionalization as 0.5 × ELF + 0.5 × religious fractionalization.
[11]This measure is based on three dimensions: (*a*) racial differences; (*b*) linguistic, national, or tribal differences; and (*c*) established religious communities. What Vanhanen (1999) called division is measured as the percentage of the largest group in the population. All three inverse percentages are summed to create the index.

in one region (see also Buhaug & Rød 2006, Weidmann et al. 2010). By and large, these findings are in line with Horowitz's (2000) expectation that the link between conflict and ethnic diversity would not be monotonic.

Still, other studies of diversity and conflict have sought to reorient the focus toward politically relevant cleavages and institutions, with mixed results. For example, Cederman & Girardin (2007) note that demographic measures of diversity (like fractionalization and polarization) do not capture the central role of the state in civil wars and do not build on a clearly articulated set of causal mechanisms for group mobilization. To capture government-group relations, the authors code what they call an N^* index of ethnonationalist exclusion[12] [using a preliminary list of ethnic groups in power, a predecessor to the Ethnic Power Relations (EPR) data described below] for Eurasia and North Africa and show that it raises the probability of ethnic civil war onset in those regions. Fearon et al. (2007) attribute this result to coding limitations, however, and argue that there is only weak evidence to support the notion that government control by minority ethnic groups is associated with increases in the probability of civil war onset. Meanwhile, Chandra & Wilkinson (2008) distinguish between ethnic structure and ethnic practice. They show that the ethnic imbalance in the army and the civil service at the time of independence from a colonial power (calculated for each institution as the sum of the differences between each group's share of the population and its share in that institution) and the percentage of the vote obtained by ethnic parties in a proximate election are associated with civil war onset, albeit with a limited set of observations. Wimmer et al.'s (2009) EPR data set allows them to establish that ethnopolitical configurations like exclusion, segmentation, and incohesion—rather than diversity per se—increase the likelihood of civil war onset (see also Cederman et al. 2010, Smith 2013, Wucherpfennig et al. 2012). This emphasis on the political dimensions of ethnic difference is in stark contrast to a more recent study by Arbatlı et al. (2020), which stresses the effect of interpersonal diversity (measured via a heterozygosity index, which is essentially a version of ELF, or measure 3, applied to alleles of particular genes in a population) on conflict.

As with other literatures, many of the divergences in findings in the conflict literature are driven by differences in sources of data and drawing of boundaries between groups, rather than substantial variation in mathematical formulae. While measures of polarization may seem more theoretically appropriate than measures of fractionalization for the study of conflict, scholars have made compelling arguments for using measures and developing underlying data that capture the political relevance of societal groups.

3.4. Crime

Finally, the effect of ethnic diversity on crime has been a subject of much scholarly debate. Since the early days of criminology, efforts to explain variation in crime rates have often

[12]The N^* index is computed as pr(conflict) = $1 - \prod_{i=1}^{n-1} 1 - p(i)$, where $p(i)$ is the probability of conflict between the ethnic group in power (EGIP) and the ith ($i = 1\ldots n-1$) ethnic group. $1 - p(i)$ represents the probability of no conflict between the EGIP and the ith ethnic group. N^* is obtained after substituting in a suitable function for $p(i)$. Under the assumption that dyadic conflicts between the EGIP and any other ethnic group are independent, the probability of no conflict between the EGIP and all other ethnic groups is simply the product of these disjoint pairwise probabilities. Thus, the probability that there is at least one source of conflict is the complement of the product (full details are available in Cederman & Girardin 2007, pp. 176–77).

turned to racial and ethnic heterogeneity as a possible influence. Overall, the expectation is that high levels of fractionalization result in more crime in a community. Drawing on prominent theories of crime and delinquency, scholars have investigated this relationship at many levels, from the national level down to the block or street level. However, by and large, this literature continues to follow the precedent set by Alesina et al. (2003), using the same measures and/or continuing to calculate fractionalization using HHI (measure 4) or ELF (measure 3). Interestingly, despite using similar measures, the results tend to be mixed, suggesting that the level of aggregation is critical to understanding the effects of this relationship.

The major theoretical perspectives in criminology have generally hypothesized that greater racial and ethnic heterogeneity lead to more crime. Social theories on crime tend to highlight the ways in which racial and ethnic heterogeneity create strain in societies and lead individuals to crime and deviance. McVeigh (2006) provides a concise overview of these perspectives, also described here. Going back to early social disorganization theory, Shaw & McKay (1942) argue that heterogeneity inhibits the creation of social cohesion, making it more difficult to oppose deviant behavior. Similarly, Hirschi's (1969) control theory focuses on controlling crime through shared normative values, which McVeigh (2006) argues are less likely to exist in a heterogeneous community. Reduced social cohesion is also expected when there is greater religious heterogeneity, due to potentially conflicting religious views on topics such as the abolition of the death penalty or abortion.

Another hypothesized mechanism explaining the expected relationship between crime and ethnic heterogeneity focuses on the relationship between ethnic heterogeneity and socioeconomic inequality, which in turn influences the crime rate. For example, Dancygier et al. (2022) find that antirefugee hate crime is most likely to occur in regions where native men face a disadvantage in finding female partners, particularly among those already disadvantaged in the mating market. Likewise, hate crimes are naturally uncommon in homogeneous communities, but not in heterogenous communities. Similarly, in heterogeneous communities where there is income inequality across groups, we would expect less cross-group coalition forming and greater tensions, causing feelings of alienation, disadvantage, and frustration that lead to crime. This correlates with a subjective fear of minority groups. For example, Ward (2019) finds that as an immigrant group increases its share of young men, the group is more likely to be viewed as a security threat.

In testing these theories empirically, scholars have examined the relationship between diversity and crime rates by looking at a variety of regions and levels of analysis. However, these analyses tend to be reasonably uniform in their measurement of these key predictors, relying on commonly used data and measures. Instead, level of aggregation is a more common measurement debate in this literature. For example, in their cross-national examination of the relationship between crime rates and ethnic and linguistic diversity, Churchill & Laryea (2019) examine crime rates across 78 countries. They hypothesize that country-level ethnic diversity affects social capital, in turn affecting crime rates. Similarly, fractionalization, they claim, is linked to weaker institutions, also likely leading to more crime. To test this relationship, they use the measures of ethnic diversity and linguistic diversity (measure 3) from Alesina et al. (2003); they address issues of endogeneity between these measures and the outcomes by utilizing the instrumental variables land quality, latitude, and elevation. Surprisingly, the authors find robust negative effects of both ethnic and linguistic fractionalization on crime outcomes. Specifically, they find no effect of ethnic fractionalization on assault or sexual crimes, but burglary, robbery, homicide, and theft are reduced when ethnic fractionalization is higher. Similar effects are seen for linguistic fractionalization. These results hold even when tested using alternative indices of fractionalization, such as those constructed by Alesina & Zhuravskaya (2011), which use aggregated regional data, rather than national data, in calculating national indices, providing robust support for their findings of a negative relationship between fractionalization and crime.

Closer to home, others have examined the relationship between fractionalization and crime rates domestically. For example, McVeigh (2006), using the theoretical grounding of Blau space as well as the dominant theories described above, hypothesizes that religious and ethnic heterogeneity positively affect crime. With US counties (or county equivalents) as the unit of analysis—which he argues is appropriate due to inclusion of rural and suburban areas without exclusion of intrastate variation—McVeigh calculates religious heterogeneity using data from the Churches and Church Members data set (ASARB 1990, 2000), which provides data by county of followers of different denominations and church bodies. The index is calculated using the inverse of HHI, ELF (measure 3). Racial and ethnic heterogeneity are calculated similarly using the five-category race data (white, African American, Latino, Asian, other) from the census by county to construct an HHI, of which he takes the inverse, then the natural log multiplied by 100. He also includes a measure of the percent foreign born in the county, providing an additional indicator of diversity. Both religious and ethnic fractionalization are found to have a positive effect on crime, as does the percent foreign born. Ethnic heterogeneity has a positive effect on all seven types of crime evaluated, while religious heterogeneity has a positive effect on all but the burglary rate and murder/assault rate. However, the models do a better job predicting property crime than violent crime. These results contrast with those of Churchill & Laryea (2019), suggesting that level of aggregation matters.

This concern with level of aggregation is supported by Kim (2018) in a methodological paper in which he investigates the difference between measuring crime by street segment and census block. While most of his results are the same using either unit of analysis, the most notable difference is the effect of racial/ethnic heterogeneity. Theoretically, he claims, smaller units of analysis best reflect criminal opportunity. Using 2010 census block data from three towns in Orange County, California, Kim creates three measures of diversity: (*a*) percent African American, (*b*) percent Latino/Hispanic, and (*c*) 1 – HHI based on five racial/ethnic groups. The third measure reflects that used by McVeigh (2006); however, in contrast to those results, Kim (2018) finds a negative relationship between diversity and crime at the street segment level. Yet at the census block level, a positive relationship between crime and diversity emerges. Despite high correlations between heterogeneity at both the block and street segment levels, the difference in effect is quite striking. Much of the effect is driven by changes in the effect on property crime and robbery, which Kim suggests could be a reflection of the contact hypothesis at work.

To summarize, while these articles do not vary greatly in their measurement of ethnic heterogeneity, it is clear that the unit of analysis and aggregation from which these measures are calculated or effects are observed can lead to different conclusions about the relationship between ethnic heterogeneity and crime rates. For scholars conducting research on crime and ethnic heterogeneity, ELF is typically the preferred measure; however, researchers should ensure that the level of aggregation chosen reflects their theoretical model appropriately as it can have clear consequences for their findings.

4. CONCLUSION

Throughout this review, we have highlighted the key ways in which ethnic heterogeneity has been conceptualized, measured, and utilized in the sociological literature and related political science and economics literature. While ethnic diversity has been used to explain a host of outcomes ranging from social capital to redistribution, conflict, and crime, the measurement of this concept tends to involve slight variations of similar computations. These measures are often very highly related linearly or quadratically to one another, which is unsurprising given that they are frequently specializations or generalizations of each other.

However, while measures are similar to each other mathematically, minor tweaks and applications using diverse data sources lead to divergences in the literature. Thus, the major differences in diversity terms actually emerge from variations in data sources and coding, processes which involve much more subjectivity. Researchers debate which aspects of heterogeneity should be included (e.g., linguistic or other differences between groups, power differentials between groups, or number of groups) and what should be the relevant points of division within a society. In particular, scholars debate the meaning of ethnicity, which groups matter, the best sources of data on ethnicity, and which geographic level is most appropriate for these analyses.

Based on our review of the literature, we find a few areas of emerging consensus in choice of measures and levels of measurement. There is some consensus that fractionalization is preferable when the outcome is a continuous measure of inequality. Polarization measures are preferred by some researchers for predicting outright civil conflict; because civil wars are most commonly studied, diversity in this area is generally measured at the country level. Existing research suggests that neighborhood-level ethnic diversity has the strongest relationships with both social capital and crime-related outcomes. When outcomes pertain to redistributive policies or attitudes about such policies, the level of measurement of diversity should match that of the policies of interest; no level is more widely accepted as preferable except in the case of the effects of immigrant stock and flows, in which the regional (subnational) level has consistently been shown to be most salient.

One issue with new data generation in this field is that it is often driven by theories about an outcome of particular interest to the research team. A focus on groups researchers consider politically relevant, for example, can raise issues of selection bias, endogeneity, concept validity, and replicability (Marquardt & Herrera 2015). Moreover, multidimensional measures of group similarity are more difficult to derive because, as the number of dimensions increases, group sizes become smaller. Thus, it may be difficult to derive group-based measures of similarity based on numerous characteristics. However, rigorous efforts to classify the political relevance of ethnic groups (via expert surveys, for example) can help ensure that time-variant and salient cleavages will be captured (Cederman et al. 2010). We conclude that an ideal measure of diversity would take into account not only the number of groups in a population and their relative share of the overall population but also how different groups are from one another along more than one dimension (as long as group sizes do not become too small), capturing some characteristic of power differentials between groups.

Two thus far less widely used measures that we find particularly promising include GE and GELF. While, again, it is difficult to derive group-based measures of similarity based on larger numbers of characteristics, the individualized approach illustrated by GELF (Bossert et al. 2011) can potentially represent a more nuanced degree of heterogeneity. GELF partially resolves two limitations of ELF: (*a*) its assumption that differences between groups in a population are equivalent and (*b*) its unidimensionality in terms of how groupings are commonly measured. Koopmans & Schaeffer (2015) contend that measures of both diversity and polarization aim to account for three features of population heterogeneity: the extent of variety of population subgroups, the extent of balance between the groups, and the disparities between the groups. While both types of measures try to capture all three with a single number, diversity measures emphasize variety and polarization measures emphasize balance. Thus, they developed GE, which can be reduced to measure either fractionalization or polarization while accounting for the relative importance of variety and balance for a given outcome.

Nearly all researchers contend that ethnic heterogeneity is important for social, political, and economic outcomes, and many theorize that ethnic heterogeneity will yield negative effects. Yet the results of the studies reviewed above present a more inconclusive picture. Through our discussion of the mathematical similarities between measures and differences between data sources, we hope to have provided some clarity about the crowded field of approaches to measuring ethnic diversity. Future researchers will now have a comprehensive source outlining how these measures have been used in key scholarly studies when it comes to the operationalization of ethnic heterogeneity, application of these measures, levels of measurement, and the results produced from major studies examining the effects of ethnic heterogeneity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Christel Kesler and Nate Breznau for their invaluable feedback on the first draft of this article, as well as Giovanna Jara for providing research assistance. We also acknowledge the contributions of the anonymous reviewers involved in improving our original submission.

LITERATURE CITED

Abascal M, Baldassarri D. 2015. Love thy neighbor? Ethnoracial diversity and trust reexamined. Am. J. Sociol 121(3):722–82

Alesina A, Devleeschauwer A, Easterly W, Kurlat S, Wacziarg R. 2003. Fractionalization. J. Econ. Growth 8(2):155–94

Alesina A, Murard E, Rapoport H. 2021. Immigration and preferences for redistribution in Europe. J. Econ. Geogr 21(6):925–54

Alesina A, Zhuravskaya E. 2011. Segregation and the quality of government in a cross section of countries. Am. Econ. Rev 101(5):1872–911

Annett A. 2001. Social fractionalization, political instability, and the size of government. Work. Pap. 00/82, Int. Monet. Fund, Washington, DC

Arbatlı CE, Ashraf QH, Galor O, Klemp M. 2020. Diversity and conflict. Econometrica 88(2):727–97 [PubMed: 36071951]

ASARB (Assoc. Stat. Am. Relig. Bodies). 1990. Churches and church membership in the United States, 1990. Data Set, ASARB/Lilly Endow., Lenexa, KS. 10.17605/OSF.IO/BMDAE

ASARB (Assoc. Stat. Am. Relig. Bodies). 2000. Churches and church membership in the United States, 2000. Data Set, ASARB, Lenexa, KS. 10.17605/OSF.IO/AV9KG

Baldassarri D, Abascal M. 2020. Diversity and prosocial behavior. Science 369(6508):1183–87 [PubMed: 32883860]

Baldwin K, Huber JD. 2010. Economic versus cultural differences: forms of ethnic diversity and public goods provision. Am. Political Sci. Rev 104(4):644–62

Barrett DB, ed. 1982. World Christian Encyclopedia. Oxford, UK: Oxford Univ. Press

Bossert W, D'Ambrosio C, Ferrara EL. 2011. A generalized index of fractionalization. Economica 78(312):723–50

Brady D, Finnigan R. 2014. Does immigration undermine public support for social policy? Am. Sociol. Rev 79:17–42

Breznau N, Rinke EM, Wuttke A, Nguyen HHV, Adem M, et al. 2021. How many replicators does it take to achieve reliability? Investigating researcher variability in a crowdsourced replication. SocArXiv https://osf.io/preprints/socarxiv/j7qta/

Bruk SI, Alenčenko VS, Telberg VG. 1964. Atlas Narodov Mira. Moscow: Akad. Nauk SSSR

Buhaug H, Rød JK. 2006. Local determinants of African civil wars, 1970–2001. Political Geogr. 25(3):315–35

Campos NF, Kuzeyev VS. 2007. On the dynamics of ethnic fractionalization. Am. J. Political Sci 51(3):620–39

Cederman L-E, Girardin L. 2007. Beyond fractionalization: mapping ethnicity onto nationalist insurgencies. Am. Political Sci. Rev 101(1):173–85

Cederman L-E, Wimmer A, Min B. 2010. Why do ethnic groups rebel? New data and analysis. World Politics 62(1):87–119

Chandra K, Wilkinson S. 2008. Measuring the effect of "ethnicity." Comp. Political Stud 41(4/5):515–63

Churchill SA, Laryea E. 2019. Crime and ethnic diversity: cross-country evidence. Crime Delinq. 65(2):239–69

Collier P, Hoeffler A. 2004. Greed and grievance in civil war. Oxf. Econ. Pap 56(4):563–95

Collier P, Hoeffler A, Rohner D. 2008. Beyond greed and grievance: feasibility and civil war. Oxf. Econ. Pap 61(1):1–27

Dancygier R, Naoki E, Amaney J, Rischke R. 2022. Hate crimes and gender imbalances: fears over mate competition and violence against refugees. Am. J. Political Sci 66(2):501–15

Desmet K, Ortuño-Ortín I, Wacziarg R. 2012. The political economy of linguistic cleavages. J. Dev. Econ 97(2):322–38

Desmet K, Ortuño-Ortín I, Weber S. 2005. Peripheral diversity and redistribution. Discuss. Pap. 2005/44, CORE, UC Louvain, Louvain, Belg.

Desmet K, Ortuño-Ortin I, Weber S. 2009. Linguistic diversity and redistribution. J. Eur. Econ. Assoc 7(6):1291–318

Dinesen PT, Schaeffer M, Sønderskov KM. 2020. Ethnic diversity and social trust: a narrative and meta-analytical review. Annu. Rev. Political Sci 23:441–65

Dinesen PT, Sønderskov KM. 2015. Ethnic diversity and social trust: evidence from the micro-context. Am. Sociol. Rev 80(3):550–73

Drazanova L. 2020. Introducing the Historical Index of Ethnic Fractionalization (HIEF) dataset: accounting for longitudinal changes in ethnic diversity. J. Open Humanit. Data 6(1):6

Duclos J-Y, Esteban J, Ray D. 2004. Polarization: concepts, measurement, estimation. Econometrica 72(6):1737–72

Easterly W, Levine R. 1997. Africa's growth tragedy: policies and ethnic divisions. Q. J. Econ 112(4):1203–50

Eger MA, Breznau N. 2017. Immigration and the welfare state: a cross-regional analysis of European welfare attitudes. Int. J. Comp. Sociol 58(5):440–63

Elbadawi E, Sambanis N. 2000. Why are there so many civil wars in Africa? Understanding and preventing violent conflict. J. Afr. Econ 9(3):244–69

Ellingsen T. 2000. Colorful community or ethnic witches' brew? Multiethnicity and domestic conflict during and after the Cold War. J. Confl. Resolut 44(2):228–49

Esteban J, Mayoral L, Ray D. 2012. Ethnicity and conflict: an empirical study. Am. Econ. Rev 102(4):1310–42

Esteban J, Ray D. 1994. On the measurement of polarization. Econometrica 62(4):819–51

Esteban J, Ray D. 2011. Linking conflict to inequality and polarization. Am. Econ. Rev 101(4):1345–74

Eur. Values Study Group, World Values Surv. Assoc. 2006. European and World Values Surveys four-wave integrated data file, 1980–2004. Data File, ZA5080 version 1.0, ZA, Cologne, Ger. 10.4232/1.11159

Fearon JD. 2003. Ethnic and cultural diversity by country. J. Econ. Growth 8(2):195–222

Fearon JD, Kasara K, Laitin DD. 2007. Ethnic minority rule and civil war onset. Am. Political Sci. Rev 101(1):187–93

Fearon JD, Laitin DD. 2003. Ethnicity, insurgency, and civil war. Am. Political Sci. Rev 97(1):75–90

Finseraas H. 2012. Poverty, ethnic minorities among the poor, and preferences for redistribution in European regions. J. Eur. Soc. Policy 22(2):164–80

Forsberg E. 2008. Polarization and ethnic conflict in a widened strategic setting. J. Peace Res 45(2):283–300

Greenberg JH. 1956. The measurement of linguistic diversity. Language 32(1):109–15

Gurr TR. 2009. MAR data. Minorities at Risk Project Data Set, Cent. Int. Dev. Confl. Manag., Univ. Md., College Park. http://www.mar.umd.edu/mar_data.asp

Habyarimana J, Humphreys M, Posner DN, Weinstein JM. 2007. Why does ethnic diversity undermine public goods provision? Am. Political Sci. Rev 101(4):709–25

Hainmueller J, Hiscox MJ. 2010. Attitudes toward highly skilled and low-skilled immigration: evidence from a survey experiment. Am. Political Sci. Rev 104(1):61–84

Hegre H, Sambanis N. 2006. Sensitivity analysis of empirical results on civil war onset. J. Confl. Resolut 50(4):508–35

Hirschi T. 1969. Causes of Delinquency. Berkeley: Univ. Calif. Press

Horowitz DL. 2000. Ethnic Groups in Conflict. Berkeley: Univ. Calif. Press. 2nd ed.

Kesler C, Bloemraad I. 2010. Does immigration erode social capital? The conditional effects of immigration-generated diversity on trust, membership, and participation across 19 countries, 1981–2000. Can. J. Political Sci 43(2):319–47

Kim Y-A. 2018. Examining the relationship between the structural characteristics of place and crime by imputing census block data in street segments: Is the pain worth the gain? J. Quant. Criminol 34(1):67–110

Koopmans R, Schaeffer M. 2015. Relational diversity and neighbourhood cohesion: unpacking variety, balance and in-group size. Soc. Sci. Res 53:162–76 [PubMed: 26188445]

Lieberman ES, Singh P. 2012. Conceptualizing and measuring ethnic politics: an institutional complement to demographic, behavioral, and cognitive approaches. Stud. Comp. Int. Dev 47(3):255–86

Marquardt KL, Herrera YM. 2015. Ethnicity as a variable: an assessment of measures and data sets of ethnicity and related identities. Soc. Sci. Q 96(3):689–716

Mau S, Burkhardt C. 2009. Migration and welfare state solidarity in Western Europe. J. Eur. Soc. Policy 19(3):213–29

McVeigh R. 2006. Structural influences on activism and crime: identifying the social structure of discontent. Am. J. Sociol 112(2):510–66

Montalvo JG, Reynal-Querol M. 2005. Ethnic polarization, potential conflict, and civil wars. Am. Econ. Rev 95(3):796–816

Mozaffar S, Scarritt JR, Galaich G. 2003. Electoral institutions, ethnopolitical cleavages, and party systems in Africa’s emerging democracies. Am. Political Sci. Rev 97(3):379–90

Okediji TO. 2005. The dynamics of ethnic fragmentation. Am. J. Econ. Sociol 64(2):637–62

Østby G. 2008. Polarization, horizontal inequalities and violent civil conflict. J. Peace Res 45:143–62

Patsiurko N, Campbell JL, Hall JA. 2012. Measuring cultural diversity: ethnic, linguistic and religious fractionalization in the OECD. Ethn. Racial Stud 35(2):195–217

Portes A, Vickstrom E. 2011. Diversity, social capital, and cohesion. Annu. Rev. Sociol 37:461–79

Posner DN. 2000. Ethnic fractionalization in Africa: How should it be measured? What does it explain about economic growth? Paper presented at World Bank Development Research Group Seminar, Washington, DC, March 29

Posner DN. 2004. Measuring ethnic fractionalization in Africa. Am. J. Political Sci 48(4):849–63

Putnam RD. 2007. E pluribus unum: diversity and community in the twenty-first century. Scand. Political Stud 30(2):137–74

Ramos MR, Bennett MR, Massey DS, Hewstone M. 2019. Humans adapt to social diversity over time. PNAS 116(25):12244–49 [PubMed: 31072924]

Reeskens T, Van Oorschot W. 2012. Disentangling the “new liberal dilemma”: on the relation between general welfare redistribution preferences and welfare chauvinism. Int. J. Comp. Sociol 53(2):120–39

Reynal-Querol M. 2002. Ethnicity, political systems, and civil wars. J. Confl. Resolut 46(1):29–54

Scheepers P, Gijsberts M, Coenders M. 2002. Ethnic exclusionism in European countries: public opposition to civil rights for legal migrants as a response to perceived ethnic threat. Eur. Sociol. Rev 18(1):17–34

Schmidt-Catran AW, Spies DC. 2016. Immigration and welfare support in Germany. Am. Sociol. Rev 81(2):242–61

Semyonov M, Raijman R, Gorodzeisky A. 2008. Foreigners’ impact on European societies: public views and perceptions in a cross-national comparative perspective. Int. J. Comp. Sociol 49(1):5–29

Senik C, Stichnoth H, Van der Straeten K. 2009. Immigration and natives’ attitudes towards the welfare state: evidence from the European Social Survey. Soc. Indic. Res 91(3):345–70

Shaw CR, McKay HD. 1942. Juvenile Delinquency and Urban Areas. Chicago: Univ. Chicago Press

Smith B. 2013. Separatist conflict in the former Soviet Union and beyond: How different was communism? World Politics 65(2):350–81

Steele LG. 2016. Ethnic diversity and support for redistributive social policies. Soc. Forces 94(4):1439–81

Steele LG, Abdelaaty L. 2019. Ethnic diversity and attitudes towards refugees. J. Ethn. Migr. Stud 45(11):1833–56

Steele LG, Perkins KM. 2019. The effects of perceived neighborhood immigrant population size on preferences for redistribution in New York City: a pilot study. Front. Sociol 4:18 [PubMed: 33869344]

Stichnoth H, Van der Straeten K. 2013. Ethnic diversity, public spending, and individual support for the welfare state: a review of the empirical literature. J. Econ. Surv 27(2):364–89

Toft MD. 2005. The Geography of Ethnic Violence. Princeton, NJ: Princeton Univ. Press

United Nations. 2005. Trends in total migrant stock: the 2005 revision. CD-ROM Doc., United Nations, New York. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Jan/un_2005_internationalmigrationstocks.pdf

Van der Meer T, Tolsma J. 2014. Ethnic diversity and its effects on social cohesion. Annu. Rev. Sociol 40:459–78

Vanhanen T. 1999. Domestic ethnic conflict and ethnic nepotism: a comparative analysis. J. Peace Res 36(1):55–73

Ward DG. 2019. Public attitudes toward young immigrant men. Am. Political Sci. Rev 113(1):264–69

Weidmann NB, Rød JK, Cederman L-E. 2010. Representing ethnic groups in space: a new dataset. J. Peace Res 47(4):491–99

Wimmer A, Cederman L-E, Min B. 2009. Ethnic politics and armed conflict: a configurational analysis of a new global data set. Am. Sociol. Rev 74(2):316–37

Wong D. 2009. The modifiable areal unit problem (MAUP). In The SAGE Handbook of Spatial Analysis, eds. Fotheringham AS, Rogerson P, pp. 105–24. Los Angeles, CA: SAGE

Wucherpfennig J, Metternich NW, Cederman L-E, Gleditsch KS. 2012. Ethnicity, the state, and the duration of civil war. World Politics 64(1):79–115

Table 1

Prominent measures of diversity and polarization[a]

Name	Equation	Comments	Users/developers
1. PP	$\sum_{j=1}^{G} p_j$	p_j is % in group j. It is maximized when $p_j = p_k$.	None reviewed
2. Gini coefficient or BGI	$(1/2\bar{\bar{y}}) \sum_{j=1}^{G} \sum_{k=1}^{G} p_j p_k \lvert \bar{y}_j - \bar{y}_k \rvert$	$\bar{\bar{y}}$ is the grand mean of y (weighted by group size). Formula shown is for groups, but can be individuals if $n = 1$ in each group. Numerous alternative but equivalent calculations exist, most at individual rather than group level.	Baldwin & Huber 2010, Esteban & Ray 2011, Esteban et al. 2012, McVeigh 2006
3. ELF	$1 - \sum_{j=1}^{G} p_j^2 \equiv \sum_{j=1}^{G} p_j(1 - p_j)$	This is the most common measure used in the reviewed literature. Groups may be defined other than via ethnicity or language.	Abascal & Baldassarri 2015; Alesina & Zhuravskaya 2011; Alesina et al. 2003; Annett 2001; Baldwin & Huber 2010; Campos & Kuzeyev 2007; Churchill & Laryea 2019; Collier & Hoeffler 2004;[b]Desmet et al. 2009, 2012; Dinesen & Sønderskov 2015; Drazanova 2020; Easterly & Levine 1997; Esteban & Ray 2011; Esteban et al. 2012; Fearon 2003, 2007; Kim 2018;Mau & Burkhardt 2009; McVeigh 2006; Montalvo & Reynal-Querol 2005; Mozaffar et al. 2003;[b]Okediji 2005;[b]Patsiurko et al. 2012; Posner 2000;[b]Putnam 2007; Ramos et al. 2019; Steele 2016; Steele & Abdelaaty 2019; Wimmer et al. 2009
4. HHI	$\sum_{j=1}^{G} p_j^2$	Measure of concentration Complement of ELF	See references above: Some call ELF the HHI.
5. CF or quadratic entropy	$1 - \sum_{j=1}^{G} \sum_{k=1}^{G} p_j p_k s_{jk}$	Calculated at group level s_{jk} is a linguistic similarity measure of groups on the [0,1] interval. Equivalent to ELF if s_{jk} is limited to 0 or 1	Baldwin & Huber 2010, Desmet et al. 2009, Fearon 2003, Steele 2016, Steele & Abdelaaty 2019
6. GELF	$1 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} s_{ij}$	This is the same as CF but calculated at the individual level. s_{ij} is any similarity measure one can compute scaled to [0, 1] interval.	Bossert et al. 2011
7. ER polarization	$C \sum_{j=1}^{G} \sum_{k=1}^{G} p_j^{1+\alpha} p_k d_{jk}$	α a is an affinity parameter, emphasizing within-group cohesion if $\alpha > 0$. d_{jk} are distance, rather than similarity, measures. This quantity is proportional to CF if $\alpha = 0$ and to HHI if $d_{jk} = I(j = k)$.	Collier & Hoeffler 2004, Desmet et al. 2009, Duclos et al. 2004,[b]Esteban & Ray 1994, Esteban et al. 2012, Montalvo & Reynal-Querol 2005, Østby 2008, Steele 2016, Steele & Abdelaaty 2019
8. RQ polarization	$1 - \sum_{j=1}^{G} \left(\frac{1/2 - p_j}{1/2}\right)^2 p_j \equiv 4 \sum_{j=1}^{G} p_j^2 (1 - p_j)$	RQ is similar to ER but centers polarization over a maximum of 50/50 population split. It assumes $\alpha = 1$ and ignores distance.	Desmet et al. 2009, 2012; Østby 2008; Reynal-Querol 2002; Steele & Abdelaaty 2019
9. Peripheral diversity or	$\sum_{j=1}^{G-1} \left[p_j^{1+\alpha} d_{0j} + p_j p_0^{1+\alpha} d_{0j} \right]$	This measures diversity/ polarization only from the dominant group ($j = 0$).	Desmet et al. 2005, Østby 2008, Steele & Abdelaaty 2019

Name	Equation	Comments	Users/developers
polarization (α > 0)			
10. GE	$\left(\sum_{j=1}^{G} p_j^{\alpha}\right)^{\frac{1}{1-\alpha}}$	This attempts to capture both variety and balance with α a in two positions.	Koopmans & Schaeffer 2015

[a]To keep notation consistent, we use the following: n is a number of individuals, whereas G is used to represent the number of groups in a population. p is used to represent the proportion of the population in a particular group. s_{jk}/d_{jk} is used to represent a similarity/difference measure on the [0,1] interval between groups or individuals j and k, respectively. C is used as a general constant, j and k are used as indices for groups, and i is used as an index for individuals. When j is used in conjunction with i (as in GELF), it also refers to individuals. Finally α is used as an affinity parameter and is the key parameter that generally differentiates polarization measures from diversity measures.

[b]Slight deviation from the original measure. For example, Mozaffar et al. (2003) use politically relevant ethnic groups, as developed by Posner (2000) and represented by Posner (2004) in the literature reviewed herein. Okediji (2005) uses a three-way version of ELF. Duclos et al. (2004) extend the original ER polarization measure (Esteban & Ray 1994) to handle continuous distances.

Abbreviations: BGI, between-group inequality; CF, cultural fractionalization; ER, Esteban-Ray; ELF, ethno-linguistic fraction; GE, generalized entropy; GELF, generalized ethno-linguistic fractionalization; HHI, Herfindahl-Hirschman index; PP, product-of-proportions; RQ, Reynal-Querol.