Psychological Methods
2012, Vol. 17, No. 2, 215–227

1082-989X/12/$12.00 DOI: 10.1037/a0027129

How to Measure Diversity When You Must

David V. Budescu
Fordham University

Mia Budescu
Temple University

Racial/ethnic diversity has become an increasingly important variable in the social sciences. Research from multiple disciplines consistently demonstrates the tremendous impact of ethnic diversity on individuals and organizations. Investigators use a variety of measures, and their choices can affect the conclusions that can be drawn and limit the ability to compare and generalize results across studies effectively. The current article reviews 3 popular approaches to the measurement of diversity: the simplistic majority–minority approach and 2 multiple categories variants, the generalized variance and the lesser used entropy statistic. We discuss the properties of each approach and reject the majority–minority approach. We provide 5 examples using the generalized variance and entropy statistics and illustrate their versatility and flexibility. We urge investigators to adopt these multicategory measures and to use our discussion to determine which measure of diversity is most appropriate given the nature of one's data set and research question.

Keywords: diversity, entropy, generalized variance, multinomial distribution, variety diversity

It is often said that diversity is the "spice of life," a sentiment that is clearly reflected in research from multiple subfields of psychology. Traditionally, diversity is thought of as "the position of a population along a continuum ranging from homogeneity to heterogeneity with respect to one or more qualitative variables" (Lieberson, 1969, p. 851). However, a recent review by Harrison and Klein (2007) identified at least three types of diversities in organizations. *Separation diversity* refers to differences or disagreements on attitudes or opinions among members of a population (e.g., disagreements along ideological lines among political parties), and it is measured by regular measures of scatter (such as variance, mean absolute difference) and/or distance. *Disparity diversity* refers to dispersion along a hierarchical continuum within a particular setting (such as differences in pay, benefits, wealth, status, or power), and it is measured either by relative measures of dispersion, such as the Coefficient of Variation, or by measures of inequality, such as Gini's coefficient. Finally, *variety diversity* fits Lieberson's (1969) traditional definition, as it captures differences in group composition in a population on some categorical variable (i.e., race, religion, eye color, etc.).

The current article focuses exclusively on the measurement of the latter, variety diversity, a topic that has not received the proper attention in the psychological literature. Although all our results can be applied to measurement of diversity along any categorical variable, we will focus on racial/ethnic diversity because of the popularity and prominence of this attribute in current research. Despite the important role diversity plays in empirical research, there has been a relative dearth of literature discussing its measurement, with even less discussion of the implications of one's choice of measure. The current article summarizes various measures of diversity, discusses their interpretation, and compares their merits and shortcomings. In addition, it highlights the use of these measures in a variety of contexts and research applications.

We should clarify that by using racial/ethnic diversity as a prototype for variety diversity, we do not intend to essentialize the meaning of race. We recognize that racial/ethnic diversity is a complex construct, which could also be conceptualized as separation or disparity diversity. However, much of the psychological literature on diversity has dealt with race exclusively as a membership variable (Bernell, Mijanovich, & Weitzman, 2009; Hunt, Wise, Jipguep, Cozier, & Rosenberg, 2007; Juvonen, Nishina, & Graham, 2006; Richard, Murthi, & Ismail, 2007; Seaton & Yip, 2009). Although the current article encourages researchers to think critically about their choice of measurement when describing demographic diversity, we also urge psychologists to think about the meaning of race (and other socially constructed categories) in other than simply categorical terms.

Diversity in Psychological Research

The critical role of racial diversity in schools, neighborhoods, and work settings has received ample empirical support (see Bernell et al., 2009; Gottfredson et al., 2008; Hunt et al., 2007; Juvonen et al., 2006; Richard et al., 2007; Seaton & Yip, 2009). Research, as well as public debates on racial diversity, has contributed to important public policy changes such as desegregation and more recently affirmative action. For instance, research shows that residing and interacting in ethnically diverse settings is related to a decrease in prejudicial behavior and attitudes and increased

This article was published Online First February 6, 2012.

David V. Budescu, Department of Psychology, Fordham University, and Mia Budescu, Department of Psychology, Temple University.

The two authors contributed equally to this article. Order of authorship was determined by a coin toss. David V. Budescu's work was supported in part by National Science Foundation Grant SES-1049208. We wish to thank Tzur Karelitz, Emily Tang, Jay Verkuilen, and Tiffany Yip for many useful comments on an earlier version of the article.

Correspondence concerning this article should be addressed to David V. Budescu, Department of Psychology, Fordham University, 220 Dealy Hall, 441 East Fordham Road, Bronx, NY 10458. E-mail: budescu@fordham.edu

liking toward other group members (Allport, 1954; Pettigrew & Tropp, 2006; Tajfel, 1982; Zajonc, 1968). Similarly, growing up in a racially or ethnically diverse setting has been shown to be related to differential cognitive, social, and civic outcomes (Hurtado, 2005), and interacting with a diverse group of peers is thought to encourage more active and creative thinking among children and adults (Gurin, Dey, Gurin, & Hurtado, 2003; Piaget, 1977; Sommers, Warp, & Mahoney, 2008).

Similar effects have been documented in educational settings. Gottfredson et al. (2008) found, in a nationally representative sample of incoming law students, that attending a more diverse school was related to increased levels of cognitive openness. A recent large-scale study of an ethnically diverse sample of middle school students (Juvonen et al., 2006) found that higher classroom diversity was a powerful predictor of perceptions of safety and social satisfaction. Hallinan and Teixeira (1987) also found that adolescents in highly diverse classrooms tend to have more cross-ethnic friendships. Overall, there is agreement that student body diversity leads to better learning outcomes and better prepares youth to be well-rounded professionals who are capable of working in variety of settings (see Hurtado, 2005).

Other research suggests a more nuanced relationship between racial diversity and positive outcomes. Criminal justice research and theory suggest that neighborhood racial heterogeneity leads to a breakdown in social cohesion and consequently to higher levels of crimes such as car thefts and racial tension (Walsh & Taylor, 2007; Welch, Sigelman, Bledsoe, & Combs, 2001), and recently, Seaton and Yip (2009) found that higher classroom diversity was associated with higher levels of perceived cultural discrimination by African American adolescents. Interestingly, adolescents attending racially diverse schools may be at a higher risk for becoming obese depending on their own racial background (Bernell et al., 2009). Workplace-based investigations indicate that racial heterogeneity may be detrimental to short-term productivity, although it has a positive effect on productivity in the long term (Richard et al., 2007).

Finally, because of the powerful role racial diversity plays in many settings, other researchers have focused on racial heterogeneity as an outcome, rather than an independent variable. Research on religious congregations suggests that racially diverse worshipers are attracted to larger congregations that have been more recently founded (Dougherty & Huyser, 2008) and that are located in racially diverse urban neighborhoods and have ethnic minority clergy. Similarly, studies indicate that institutional policies in the workplace promote diversity. For instance, business schools that extended domestic partner benefits to same-sex couples and upheld other nondiscriminatory policies also had more ethnically diverse faculty (Cook & Glass, 2008).

This short selective review illustrates that the ecological context plays a prominent role in the psychological sciences. Unfortunately, readers are limited in their interpretation of investigations relying on diversity as a key variable due to the lack of uniformity or standardization in its measurement. There are several problems in measuring diversity. The first is, simply, that researchers use different numbers of categories to define the target attribute. The second has to do with the choice of the measures to capture diversity. Whereas some have used simple measures, such as the proportion of non-African Americans (e.g., Welch et al., 2001), others have relied on more sophisticated measures that are sensitive to the relative proportion of each ethnic or racial group to the overall composition in a particular context (e.g., Juvonen et al., 2006; Seaton & Yip, 2009). Due to the widespread use of diversity as both an independent and dependent variable, it is necessary to examine carefully its measurement and understand the implications of the heterogeneity of measures across articles. Most importantly, researchers should be sensitive to the possibility that the choice of a measure may affect the conclusions that can be drawn from any one study and/or limit their ability to compare results across studies meaningfully.

Measures of Diversity

As indicated in the introduction, a measure of (variety) diversity seeks to locate any given population along a continuum ranging from homogeneity to heterogeneity with respect to one or more qualitative variables (Lieberson, 1969). Assume that the target categorical variable can take C distinct values (categories). Let P_i be the proportion of cases in category i, where $i = 1 \ldots C$, all $P_i \geq 0$, and $\sum_{i=1}^{C} P_i = 1$. The C (mutually exclusive and exhaustive) proportions define a multinomial distribution (e.g., Wickens, 1989). A measure of (variety) diversity, D, is a single-valued function of this distribution, that is, $D = f(P_1 P_2 \ldots P_C)$, that satisfies the following desiderata:

- It is bounded from above and below.
- Its minimal value is obtained when all the observations are concentrated in one category (e.g., $P_1 = 1$ and $P_2 = \ldots = P_C = 0$), that is, no diversity.
- Its maximal value is obtained when all C categories are equally represented ($P_1 = P_2 = \ldots = P_C = 1/C$), that is, maximal diversity.
- Given that the C categories refer to a categorical (nominal) variable, the measure is invariant across all transformations that preserve the identity and integrity of the C categories.

Next we review several measures of diversity.

The Simplistic Majority–Minority Approach

Diversity can be measured by focusing, simply, on the proportion of individuals belonging to a particular (majority or minority) group in a given context. In the context of race in the United States, this approach most often translates to calculating the percentage of individuals who are, or are not, White. Formally, if P_W is the proportion of Whites, then $P_{NW} = (1 - P_W)$ is the proportion of non-Whites. This approach is used in many articles (e.g., Bernell et al., 2009; Cook & Glass, 2008; Gurin, Dey, Gurin, & Hurtado, 2003; Hallinan & Teixeira, 1987; Hunt et al., 2007; Hurtado, 2005; Sommers et al., 2008; Welch et al., 2001), and it may be sufficient in some situations, such as the (very) special case in which there are only two groups ($C = 2$), but for the most part if fails to capture the full meaning of diversity. As research has consistently shown, the presence of other ethnic or racial groups has a significant impact on the relationship between majority and minority group members. Consider, for example, the study of neighborhood racial segregation on experiences of discrimination by African American residents. Census tracts that are predominantly Black experience less racial tensions than blocks that had a lower percentage of Black residents (Hunt et al., 2007). However, this measure of diversity is insensitive to the ethnic background of the non-Black residents of the census tract. A census tract that is only 25% African American, for instance, may be 75% White, or it may comprise an equal proportion of three or four distinct racial groups. Importantly, this approach focuses on the experience of the particular racial/ethnic group of interest, while homogenizing the outgroup members.

A slightly superior, yet equally coarse approach, often used by criminologists and sociologists to measure neighborhood heterogeneity (see Miethe & McDowall, 1993; Rice & Csmith, 2002; Walsh & Taylor, 2007), defines diversity as the product of the two proportions, that is, $P_W P_{NW} = P_W(1 - P_W)$. This is immediately recognized as the variance of a *binomial distribution* implied by the majority–minority distinction. It has the obvious advantage that it peaks (.25) when the diversity is maximal (when $P_W = P_{NW} = .5$), and it approaches 0 as the group becomes more homogeneous, that is, as P_W (or P_{NW}) approaches 1.0, and satisfies the four desiderata. However, it suffers from the same glaring weakness as P_W (or P_{NW}), namely, insensitivity to the distribution of the non-White ethnic or racial groups that compose the population.

Alternative Approaches: Generalized Variance

A more satisfactory and complete measure of diversity can be obtained by considering the full distribution over the *C* categories. As indicated above, the *C* proportions are jointly distributed as a (multivariate) multinomial distribution (e.g., Wickens, 1989). The variance of each proportion is $P_i(1 - P_i)$, and the covariance between any pair of distinct proportions, *i* and *i*′, is $-P_i P_i$ (when $C = 2$, the multinomial is the regular binomial). It is customary in multivariate analysis to consider generalized measures of variance—single numerical values that summarize the variability of the whole system. The two most popular measures of generalized variance (GV) are the determinant and the trace of the covariance matrix (e.g., Johnson & Wichern, 2002). The latter—the sum of the variances of the *C* categories—is often used as a measure of diversity because of its simple and easy-to-calculate, and easy-to-interpret, form:

$$\mathrm{GV} = \sum_{i=1}^{C} P_i(1 - P_i) = 1 - \sum_{i=1}^{C} P_i^2.$$

Haberman (1982) referred to this measure as the "concentration index" and traced its origins to Gini (1912). This measure was repeatedly reinvented and used in various disciplines under different names (e.g., in linguistics, Bachi, 1956; in sociology, Blau, 1977; in finance, Hirshman, 1964; in biology, Simpson, 1949). It has also been used in recent publications in psychology and related fields (see Gottfredson et al., 2008; Juvonen et al., 2006; Richard et al., 2007; Seaton & Yip, 2009). In addition to its standard statistical interpretation as the trace of the covariance matrix of the multinomial distribution, GV has a simple and intuitive interpretation: It is the probability that two randomly selected individuals from a particular population belong to different subgroups (e.g., Kader & Perry, 2007).[1] A higher value (probability) reflects a higher degree of diversity. For example, a particular neighborhood that consists of 25% African Americans, 33% Hispanics, and 42% Whites has a GV index of 0.65, and a neighborhood with 55% Whites, 25% African Americans, and 20% Hispanics has a GV index of 0.40; so the former is more diverse.

Some properties of the GV are listed below:

1. GV is invariant under any permutation of the categories and their relabeling. In other words, two identical distributions share the same GV regardless of the labeling of groups. For example, if Population 1 consists of $P(\text{White}) = .50$, $P(\text{Black}) = .25$, and $P(\text{Hispanic}) = .25$ and Population 2 of $P(\text{White}) = .25$, $P(\text{Asian}) = .25$, and $P(\text{Hispanic}) = .50$, then GV(Population 1) = GV(Population 2) $= (1 - .5^2 - .25^2 - .25^2) = .625$.

2. GV is bounded from below and above: Its minimal value is 0, and it is achieved when all the observations are concentrated in one category (e.g., $P_1 = 1$ and $P_2 = \ldots P_C = 0$); its maximal value is $(C - 1)/C$, and it is achieved when the distribution over the *C* groups is uniform (i.e., $P_1 = P_2 = \ldots = P_C = 1/C$).

3. Property 2 suggests that this diversity index is sensitive to the number of categories (groups) in a particular setting. To alleviate problems that can arise when seeking to compare distributions that have different numbers of categories, one can "normalize" GV (see Agresti & Agresti, 1978) relative to its upper bound. For any number of groups, *C*, let the normalized GV (NGV) be

$$\mathrm{NGV} = \mathrm{GV}/\mathrm{Max}(\mathrm{GV}) = \frac{C}{(C-1)}\left(1 - \sum_{i=1}^{C} P_i^2\right).$$

NGV is a bounded ratio ($0 \le \mathrm{NGV} \le 1$) that indicates how close a GV is to its maximal possible value (upper bound), which depends on *C* and, as such, is a *relative* measure of diversity that allows direct comparisons of results from studies with different numbers of categories.

4. If two (or more) nonempty categories are combined into one, GV will inevitably decrease. For example, it is easy to see that if we combine groups *i* and *i*′ into one category, $\mathrm{GV}(C) - \mathrm{GV}(C - 1) = 2P_i P_{i'}$, which is, by definition, nonnegative. However, this does not necessarily hold for NGV. If we combine groups *i* and *i*′ into one category, we find

$$\mathrm{NGV}(C) - \mathrm{NGV}(C-1) = \frac{2(C-1)^2\, P_i P_{i'} - \mathrm{GV}(C)}{(C-1)(C-2)}.$$

This sign of this quantity is a function of the relative magnitude of the categories being combined. For example, if $P_1 = .4$, $P_2 = .3$, $P_3 = .2$, and $P_4 = .1$, we obtain GV(4 Categories) = 0.7 and NGV(4 Categories) = 0.933. If we combine the first two (large) categories, we obtain GV(3 Categories) = 0.46 and NGV(3 Categories) = 0.69, so both GV and NGV are reduced. However, when we combine small categories—more specifically, if $P_i P_{i'} < \mathrm{GVC}/2(C - 1)^2$—NGV actually increases. For example, if we combine Categories 3 and 4, we obtain GV(3 Categories) = 0.66 and NGV(3 Categories) = 0.99.

5. Since the uniform distribution defines one bound of GV (and NGV), it is reasonable to relate these measures with standard tests

[1] Strictly speaking, the sampling is with replacement from infinite populations. Therefore, if the samples used to estimate GV are small, finite sampling corrections should be applied (see Biemann & Kearney, 2010, for details).

of uniformity. To test the uniformity null hypothesis, H_0: $P_1 = P_2 = \ldots = P_C = 1/C$, one could use the Pearson chi-square goodness-of-fit test with $(C - 1)$ degrees of freedom:

$$\chi^2 = \frac{\sum_{i=1}^{C}\left(P_i - \frac{1}{C}\right)^2}{\frac{1}{C}} = C\sum_{i=1}^{C} P_i^2 - 1.$$

There is a simple inverse relationship between NGV and the ratio of this test statistic and its degrees of freedom: $(1 - \text{NGV}) = \chi^2/(C - 1)$.

6. Johnston, Berry, and Mielke (2006) proposed measures of effect size for tests of goodness of fit, ES(χ^2). It is easy to show that NGV $= -$ES(χ^2). The sign reversal is due to the fact that the test uses maximal diversity (NGV $= 1$) as the baseline (null hypothesis), so it is a test of lack of diversity. In other words, NGV can be interpreted as an effect size of departure from maximal diversity.

7. NGV is an inverse linear transform of the variance of the C observed proportions: NGV $= 1 - C\sigma^2_{P_i}$.

8. Biswas and Mandal (2010) have proposed higher order variants of this measure capturing the probability that three, four, etc., randomly selected individuals belong to different subgroups.

Alternative Approaches: Entropy

Entropy is a measure of disorder or unpredictability in a physical system, which was appropriated by Shannon (1948) when he developed information theory and by Teachman (1980) to measure diversity. It is defined as a weighted sum of the probabilities where the weights are their logarithms (typically, Base 2)[2] and assuming Log(0) = 0:

$$\text{H} = -\sum_{i=1}^{C} P_i \text{Log}_2(P_i).$$

Entropy shares many of the properties of GV: It is invariant under permutation of the categories and their relabeling, and it is double bounded. Its minimal value, 0, is obtained when all the observations are concentrated in one category (e.g., $P_1 = 1$ and $P_2 = \ldots P_C = 0$). Its maximal value is $-\text{Log}_2(1/C) = \text{Log}_2(C)$, and it is achieved when the distribution over the C groups is uniform (i.e., $P_1 = P_2 = \ldots P_C = 1/C$). In the context of quantification of racial diversity, entropy was used by Dougherty and Huyser (2008).

Entropy is also sensitive to the number of categories (groups) in a particular setting, but one can normalize entropy (NH), such that $0 \leq \text{NH} \leq 1$, for any number of groups to obtain a relative measure of diversity (e.g., Noble & Sanchez, 1993):

$$\text{NH} = \text{H/Max(H)} = -\sum_{i=1}^{C} P_i \text{Log}_2(P_i)/\text{Log}_2(C).$$

If we reduce the number of categories by combining some of them, H is reduced. For example, if we combine groups i and i' into one category, we obtain

$$\text{H}(C) - \text{H}(C-1) = P_i[\text{Log}_2(P_i + P_{i'}) - \text{Log}_2(P_i)] + P_{i'}[\text{Log}_2(P_i + P_{i'}) - \text{Log}_2(P_{i'})],$$

which is nonnegative. This does not necessarily hold for the normalized measure. In this case the direction of the change depends on the quantity:

$$\frac{P_i\text{Log}_2(P_i) + P_{i'}\text{Log}_2(P_{i'})}{(P_i + P_{i'})\text{Log}_2(P_i + P_{i'})} - \frac{\text{Log}_2(C)}{\text{Log}_2(C-1)}.$$

This sign of this quantity is a function of the magnitude of categories being combined. Consider the example from the previous section where $P_1 = .4$, $P_2 = .3$, $P_3 = .2$, and $P_4 = .1$. We obtain H(4 Categories) = 1.846 and NH(4 Categories) = 0.933. If we combine the first two (large) categories, we obtain H(3 Categories) = 1.157 and NH(3 Categories) = 0.690, so both H and NH are reduced. However, when we combine small categories, NH increases. For example, if we combine Categories 3 and 4, we obtain H(3 Categories) = 1.571 and NH(3 Categories) = 0.991.

We pointed out earlier the relationship between NGV and Pearson's chi-square. A similar inverse linear relationship exists between the entropy measure and the likelihood ratio test, G^2, that can be used to test the uniformity hypothesis. Let

$$G^2 = -2\sum_{i=1}^{C} p_i \ln\left(\frac{p_i}{1/C}\right).$$

It is possible to show that $G^2 = 2K[\text{Log}_2(\text{C}) - \text{H}] = 2K(1 - \text{NH})$, where $K = 1.443$ is the constant that converts the logarithms from Base 2 to e. Finally, NGV is inversely related to the measure of effect size, ES(G^2), developed by Johnston et al. (2006) for the likelihood ratio tests.

A unique property of the entropy that can be useful in this context is its additivity: If some of the categories of a complete distribution are combined to create a reduced distribution, it is possible to reexpress the entropy of the original (complete) distribution as a weighted combination of the entropy of the reduced distribution and the entropy measures of the (collapsed) subdistributions. Thus, if the C groups are collapsed into $(C - K)$ groups where the overall probability in the K combined group is PK, then

$$\text{H}(P_{1,\ldots}P_C) = \text{H}(P_{1,\ldots}P_{C-K}) + P_K\text{H}(P_{C-K+1,\ldots}\ldots P_K).$$

For example, the entropy of the distribution $P_1 = .4$, $P_2 = .3$, $P_3 = .2$, and $P_4 = .1$ is H(4 Categories) = 1.846. If we combine Categories 3 and 4, we obtain a new reduced distribution with three categories, $P_1 = .4$, $P_2 = .3$, and $P_{34} = .3$, and H(3 Categories) = 1.571. Within the combined category, we have $P_{3|34} = .667$ and $P_{4|34} = .333$, and H(2 Categories) = 0.918. It is easy to verify that H(3 Categories) + P_{34} H(2 Categories) = $(1.571 + 0.3^{*}0.918) = 1.846$ = H(4 Categories).

Examples and Illustrations

Next we present several examples using NGV and NH to illustrate their richness, generality, and flexibility. These examples

[2] It is possible to calculate entropy using other bases (10, e, etc.). The results are simple linear transformations of the standard form because $\text{Log}_{10}(x) = .434\text{Log}_e(x)$, $\text{Log}_{10}(x) = .301\text{Log}_2(x)$, and $\text{Log}_e(x) = .693\text{Log2}(x)$. Typically, Base 2 is used because of its interpretability.

highlight new insights one can gain from the use of the two multicategory diversity measures in research contexts compared with measures that only reflect the distinction between the majority and the minorities.

Example 1: Ranking Schools by GV and Entropy

The two diversity measures are quite similar and share many properties. Moreover, the normalizations described earlier place NGV and NH on a scale with identical end points: 0 when the distribution is concentrated in one category and 1 when the distribution over the C groups is uniform. However, their scaling is slightly different. To illustrate the relationship between the two, we analyze the ethnic distribution in all public schools in New York City (downloaded from http://www.schooldatadirect.org). There are 283 schools with a total enrollment of 156,221 students, and the classification uses $C = 5$ groups. The ethnic distribution across 281 schools (no information is available for two of the schools) is $P_{\text{White}} = .123$, $P_{\text{Black}} = .306$, $P_{\text{Hispanic}} = .477$, $P_{\text{Asian}} = .089$, and $P_{\text{Native American}} = .004$.

Table 1 summarizes some key statistics of the two (normalized) statistics, and Figure 1 plots their joint distribution. The NGV measures are slightly higher (in 83% of the schools and, on average, by 0.068) and have higher standard deviation and higher ranges. There are a couple of schools that are exclusively Hispanic, so their diversity is 0. The most diverse school in the city (NGV = 0.930 and NH = 0.888) consists of $P_{\text{White}} = .316$, $P_{\text{Black}} = .226$, $P_{\text{Hispanic}} = .286$, $P_{\text{Asian}} = .150$, and $P_{\text{Native American}} = .023$. The plot illustrates a high level of agreement between the two measures. The plot shows (and a polynomial regression confirms) that the relation has a distinctive and significant curvilinear component, but the Pearson correlation is very high ($r = .98$). The Kendall τ_b rank-order correlation between the two measures is .90, indicating that $(1 + .90)/2 = 95\%$ of the $(283 \times 282)/2 = 39{,}903$ distinct pairs of schools are ordered similarly by the two measures. These results suggest that in most cases it would make little difference which of the two measures is being used.

Example 2: Proximity to Various Diversity Benchmarks

The previous example illustrates how the two measures can be used to rank and scale units (e.g., schools, neighborhood, organizations) according to their diversity, identify the most (least) diverse, and compare units directly (e.g., school A is more diverse than school B but less diverse than school C). Occasionally, researchers may wish to compare the various units relative to particular targets that may represent a goal that the units seek to achieve, a baseline from which they started, or some (national or regional) norm. Analyses of proximity to such targets can be used to document change and trace progress. We illustrate this approach using the New York City public schools and with two distinct targets.

The first target is the diversity of the population of all children enrolled in the public schools. As mentioned earlier, the ethnic distribution across all public schools in the city is $P_{\text{White}} = .123$, $P_{\text{Black}} = .306$, $P_{\text{Hispanic}} = .477$, $P_{\text{Asian}} = .089$, and $P_{\text{Native American}} = .004$, so H = 1.77 and GV = 0.66. The second target is the population of the city. According to the American Community Survey's 3-year population estimate for 2006–2008, the racial composition of all the city's residents is $P_{\text{White}} = .351$, $P_{\text{Black}} = .234$, $P_{\text{Hispanic}} = .275$, $P_{\text{Asian}} = .117$, and $P_{\text{Native American}} = .002$, so the city's diversity measured with entropy is H = 1.91, and with the GV it is GV = 0.73. The public schools are slightly less ethnically diverse than the city (in particular, see the relatively large difference in P_{White} between the two targets).

To measure the proximity of unit (school) i to the target, we calculate

$$r_i = \log\left(\frac{\text{Diversity}_i}{\text{Diversity}_{\text{target}}}\right),$$

where Diversity can be either GV or H. This measure, being a ratio, is independent of the number of categories and, because of the logarithmic transformation, is symmetric around 0. In this metric, schools that are more (less) diverse than the target are assigned positive (negative) values, and a value of 0 is obtained when a school matches the target. Table 2 presents the joint distribution of the log ratios of the entropy measures relative to the school and New York City benchmarks (the results based on GV are highly similar and are not presented.) The majority of schools are less diverse than the city and the schools' benchmarks, and very few schools are more diverse. Note, however, that in 11% of the cases the two measures are classified differently relative to the two benchmarks.

To measure distance of unit i to the target, we calculate

$$d_i = \log\left|\frac{\text{Diversity}_i}{\text{Diversity}_{\text{target}}}\right|.$$

This measure is also independent of the number of categories but, because of the absolute value transformation, is always nonnegative and measures how close the units are to the target without distinguishing between units that are more or less diverse than the target. The vast majority of the schools (92%) are closer (in terms of their diversity) to the school population than to the city's population (only 7% show the opposite pattern), reflecting the nature of the differences between the two targets.

Table 1
Summary Measures of the Distribution of Diversity Measures in New York City Schools (N = 281)

Diversity measure	*M*	*Mdn*	Standard deviation	IQR	Range
NGV	.58	.64	.23	.29	.93
NH	.52	.53	.21	.29	.89
NGV–NH	.06	.07	.05	.07	.25

Note. IQR = interquartile range; NGV = normalized general variance; NH = normalized entropy.

Example 3: The Effects of Intervention on Diversity

The next example illustrates how one can use the diversity measures to detect the effects of interventions (broadly defined) on the composition of a target population. We analyzed the diversity of the distribution of the 95,563 applicants to, and the 35,538 students eventually enrolled in, the nine campuses of the Univer-

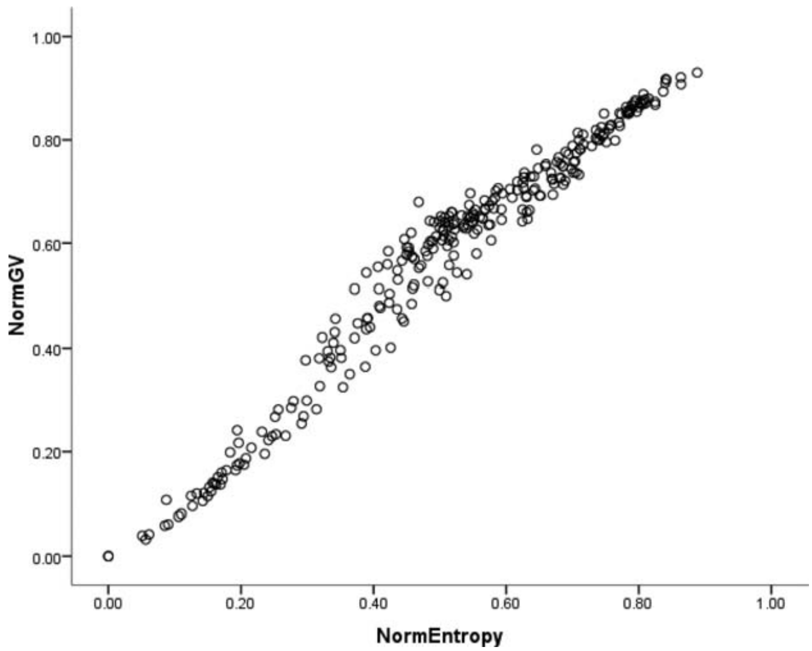

Figure 1. The joint distribution of normalized general variance (GV) and normalized entropy in all public schools in New York City (n = 283).

sity of California in the fall of 2008 (data downloaded from http://statfinder.ucop.edu/library/tables/table_3-2008.aspx). We calculated NGV and NH for the two distributions on each campus and across all of them. For the purposes of the present analysis, we excluded all foreign applicants (and enrollees), as well as those whose ethnicity was unknown, and we combined the two smallest categories (American Indian and Other) into one group. Thus, we have $C = 5$ categories: White, African American, Chicano/Latino, Asian/Filipino/Pacific Islanders, and Other/Native American.

We find that on each campus (and overall) the selection process (which reflects the combined effect of the universities' decision to admit students and the students' decisions to accept the admission offers) alters the diversity of the population. The students enrolled are less diverse than the applicants, and the standard deviation of the campus-specific measures of diversity is higher among the students enrolled than among the applicants. These patterns hold for both diversity measures. For example, when analyzing the NH, we find among applicants to the nine campuses M(NH) = 0.801 and SD(NH) = 0.024 and among the enrolled students M(NH) = 0.755 and SD(NH) = 0.055.

Although this reduction of diversity is observed on every campus, it is not uniform, as shown in Table 3, where we list the NH of the applicants and enrollees in each campus. The campuses are listed by the rate of reduction of diversity (see last column in the table). We also list, as a benchmark, the values in the University of California system as a whole. The reduction in diversity associated with the selection process is most pronounced for San Diego, Irvine, and Berkeley. In other campuses (especially Merced), it is much smaller (the results with GV are highly similar and not reported).

Table 2
Joint Distribution (%) of Diversity (Entropy) Measures Relative to the Total Population and the School Population of New York City (N = 281)

Schools compared to school population	Schools compared to city population			
	Less diverse	Equally diverse	More diverse	Total
Less diverse	86.5	0.0	0.0	86.5
Equally diverse	2.5	0.0	0.0	2.5
More diverse	7.5	1.0	2.5	11.0
Total	96.5	1.0	2.5	

Example 4: Diversity Indices as Predictors of Academic Outcomes

In the next example we demonstrate the unique contribution of the diversity indices (either entropy or the GV) as meaningful predictors of academic outcomes above and beyond demographic characteristics alone (e.g., proportion of White, Black, Hispanic, or Asian students). Thus, we illustrate how the multicategory measures of diversity contribute beyond the simplistic majority–minority measures. We use state-level data collected by the National Assessment of Educational Progress (NAEP), a project of the U.S. Department of Education that compiles periodically nationally representative data sets that include scores on a number of standardized achievement tests for fourth and eighth graders across the country (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics). The current analyses use mean scores of the 50 states and the District of Columbia on the assessments of math and reading of eighth-grade students based on assessment from the fall of 2005 (we omit the

Table 3
Measures of Diversity (Normalized Relative Entropy) of the Applicants and Enrollees on the Various Campuses of the University of California (Fall 2008)

Campus	NH(Applicants)	NH(Enrollees)	NH(Enrollees)/NH(Applicants)
San Diego	.775	.651	.840
Irvine	.779	.702	.901
Berkeley	.797	.735	.922
Santa Barbara	.804	.766	.953
Santa Cruz	.784	.749	.955
University of California system	.814	.785	.964
Los Angeles	.814	.786	.966
Riverside	.840	.814	.969
Davis	.783	.761	.972
Merced	.837	.828	.989

Note. NH = normalized entropy.

results for fourth graders that are, essentially, identical). These scores were merged with demographic information for the states (collected by the National Center for Educational Statistics in 2000) and information on the total number of schools, pupils, and teachers collected by the Common Core of Data for its 2007–2008 report (all data were downloaded from http://nces.ed.gov/nationsreportcard/).

We regressed the two outcomes (mean eighth-grade reading and math scores in each state) on nine demographic and school characteristics: the proportion of American Indian, Black, White, Hispanic, and Asian Pacific students in the state (since there is always a small number of cases with unknown race, the sum of the five proportions is, usually, slightly less than 1); the proportion of male and female residents in the state (since there is always a small number of cases with unknown gender, the sum of the two proportions is, typically, less than 1); the pupil-to-teacher ratio in the state; and one of the diversity indices. Recall that both measures are nonlinear functions of *C* proportions, so they are likely to contribute to the prediction above and beyond their constituents. Of course, we do not claim that these variables constitute the best model for predicting the NAEP scores, in a substantive sense. We use this model as a reasonable first-order approximation that allows us to illustrate the role of the diversity measures.

We conducted best subset regressions analyses for the two outcomes using the two diversity measures. We examined all possible subsets of predictors—in this case there are $(2^9 - 1) = 511$ subset models—and selected for each model size ($p = 1, 2, \ldots 9$) the one with highest R^2_{adj}. Figure 2 presents plots of the best models' fit (R^2_{adj}) for these nine best fitting models (each plot represents a different combination of the measure of diversity and the outcome variable).

The reading scores are more predictable than the math scores, and the entropy is slightly (but consistently) a better predictor than the GV for each model size. The most important point of this example, from our perspective, is the composition of the best models and their relative fits. In all four cases the single best predictor (corresponding to $p = 1$ on the *x*-axis) is the proportion of White students in the state, and the best pair of predictors (corresponding to $p = 2$ on the *x*-axis) includes the proportion of White students and the index of diversity in the state. In all models the addition of the diversity measure improves (significantly at $\alpha = .05$) the fit of the simplest model ($p = 1$), documenting the necessity of the diversity measures. This pair of predictors is included in all other best subset models (corresponding to $p = 3 \ldots 9$ in the plots), indicating their criticality (Azen, Budescu, & Reiser, 2001). Finally, note that all plots are similar in shape: Diversity increases the fit substantially, but the curves are, essentially, flat for $p > 2$. The two key predictors—proportion of White students and the index of diversity—account for, at least, 93% of the overall fit of the full models, and the seven additional variables do not contribute significantly (at $\alpha = .05$) to the model, indicating that, for all practical purposes, these two variables are sufficient.

Example 5: Decomposing Diversity to Improve Prediction of Academic Outcomes

So far we have considered global measures that reflect the diversity of a distribution over the *C* categories. In some cases, such measures are suboptimal. In particular, we are concerned by the case where the modal category dominates all others and it determines the value of the statistic to the degree that it is (almost) insensitive to the distribution over the other $(C - 1)$ categories. We illustrate this point with the NGV measure: Let $\pi = \max(P_1, P_2 \ldots P_C)$. The maximal value of NGV—obtained when the other $(C - 1)$ categories have equal probabilities, $(1 - \pi)/(C - 1)$—is

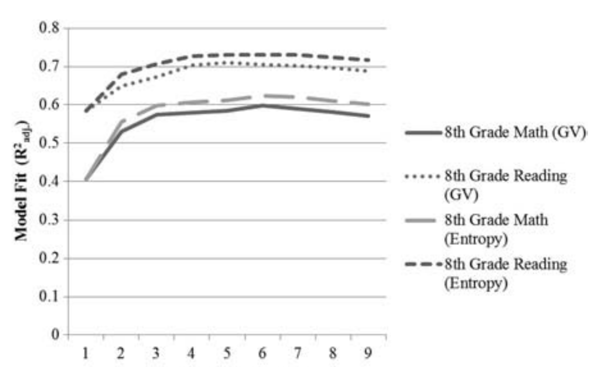

Figure 2. Adjusted R^2 values for best subsets regression of eighth-grade National Assessment of Educational Progress reading and math scores in 50 states and the District of Columbia using diversity (general variance [GV] and entropy) measures as predictors.

$$\mathrm{NGV} = \frac{C}{(C-1)}\left[1 - \pi^2 - \frac{(1-\pi)^2}{(C-1)}\right].$$

The minimal value of NGV—obtained when $(C - 2)$ categories have 0 probability and one has a probability of $(1 - \pi)$—is

$$\mathrm{NGV} = \frac{C}{(C-1)}[1 - \pi^2 - (1-\pi)^2].$$

Thus, the range of values that NGV can take (obtained by simple subtraction) is

$$\frac{C(C-2)}{(C-1)^2}(1-\pi)^2,$$

which decreases as π increases. For example, if $C = 5$ and $\pi = 0.70$, the possible range is only 0.084, and when $\pi = 0.9$, the range is reduced to 0.009!

To fully capture the diversity of the population, one can adopt a "hierarchical" approach by calculating (a) a measure of diversity that only distinguishes between the majority group and all the minorities combined into one minority group (i.e., with $C = 2$), and (b) a measure of "diversity of the minority groups" independent of the majority group (i.e., with $C - 1$ groups). We reran the regression of the NAEP scores and summarize the fit of the various models (R^2_{adj}) in Table 4. The first two data columns in the table correspond to the first two points ($p = 1$ and 2) on the curves from Example 3 in Figure 2, and in the third column we summarize models where we replaced the global diversity with the two measures described above ($p = 3$). Evidently, decomposing the global diversity into two distinct components is beneficial, as it increases (systematically, and significantly at $\alpha = .05$) the fit of the prediction models. The last column presents the fit of the best fitting model (i.e., the one with highest R^2_{adj}), which, in all cases, represents only a negligible improvement over the model with $p = 3$.

How Many Categories?

In this section we address the sensitivity of the two normalized diversity measures, NGV and NH, to changes in the number of categories defining the target distribution. We have identified earlier the circumstances under which reducing the number of categories increases or decreases the measures of relative diversity. In this section we focus on the absolute magnitude of the changes in order to determine whether they are meaningful. This also provides us with another opportunity to compare the performance of the two diversity measures.

To achieve this goal we considered all possible distributions over $C = 4$, 5, and 6 categories. To eliminate redundancies, we only consider distributions where $P_1 \leq P_2 \leq \ldots \leq P_{(C)}$. We assume that one would consider combining categories only if they are very small, so we study the cases where $P_1 = .02$ and .05. For each distribution involving C categories, we (a) calculated NGV(C) and NH(C); (b) combined the two smallest categories, P_1 and P_2, and calculated the corresponding NGV($C - 1$) and NH($C - 1$); and (c) combined the smallest category, P_1, and the largest one, P_C, and calculated the corresponding NGV($C - 1$) and NH($C - 1$).

Figures 3 and 4 depict the results for the case of $C = 4$ and $P_1 = .02$ for the two measures. They plot the Diversity($C - 1$) for the two combination schemes as a function of Diversity(C) based on all 752 relevant distributions. The reference lines represent the case where the diversity measures are unaffected by the change, that is, Diversity(C) = Diversity($C - 1$), and were added to facilitate interpretation. The lines marked "min" represent the values of Diversity($C - 1$) obtained when the two smallest categories were combined. Note that for both measures and in all these cases Diversity($C - 1$) > Diversity(C). The lines marked "max" represent the values of Diversity($C - 1$) obtained when the smallest category was combined with the largest. For both measures and in a majority of cases, Diversity($C - 1$) > Diversity(C), but for a small number of cases, where Diversity(C) is low, we observe the opposite patterns, and Diversity($C - 1$) > Diversity(C). It appears that, in general, the changes in the values of the diversity measure are not very large and that the relative entropy measure is more sensitive to the collapse of categories.

Table 5 summarizes the results for the (3 categories × 2 values of P_1) six cases investigated. For each of the original distributions, we calculated the maximal change in the relative measures of diversity across the two schemes described above. The table presents the median and the trimmed range (excluding the lowest and highest 1%) of these maximal changes. Three clear and consistent patterns emerge: (a) the effect of combining categories is most pronounced when there are few categories ($C = 4$) and is reduced as the number of original categories increases; (b) the effect of combining categories is stronger when the smallest category, P_1, is low ($P_1 = .02$) and is attenuated when $P_1 = .05$; and (c) the effects are much more pronounced for the relative entropy. This point is vividly illustrated in Figure 5, which plots the maximal changes in the two measures for the case used in the previous demonstration ($C = 4$ and $P_1 = .02$).

Table 4
Measures of Fit (R^2_{adj}) of Models Using Global and Distinct (Majority and Minority) Measures of Diversity (National Assessment of Educational Progress Scores of U.S. States and District of Columbia, N = 51)

		Model		
Outcome measure	Proportion white	Proportion white, global diversity	Proportion white, majority diversity, minority diversity	Best fitting
NGV				
Math	.41	.53	.64	.67
Reading	.58	.65	.72	.78
NH				
Math	.41	.56	.64	.67
Reading	.58	.68	.73	.77

Note. NGV = normalized general variance; NH = normalized entropy.

Summary and Recommendations

The current article is concerned with the measurement of diversity (as variety, in the Harrison & Klein, 2007, typology). We sought a single numerical value that captures the degree of (dis)similarity between the relative size of the C subpopulations defined by demographic and social categories such as ethnicity,

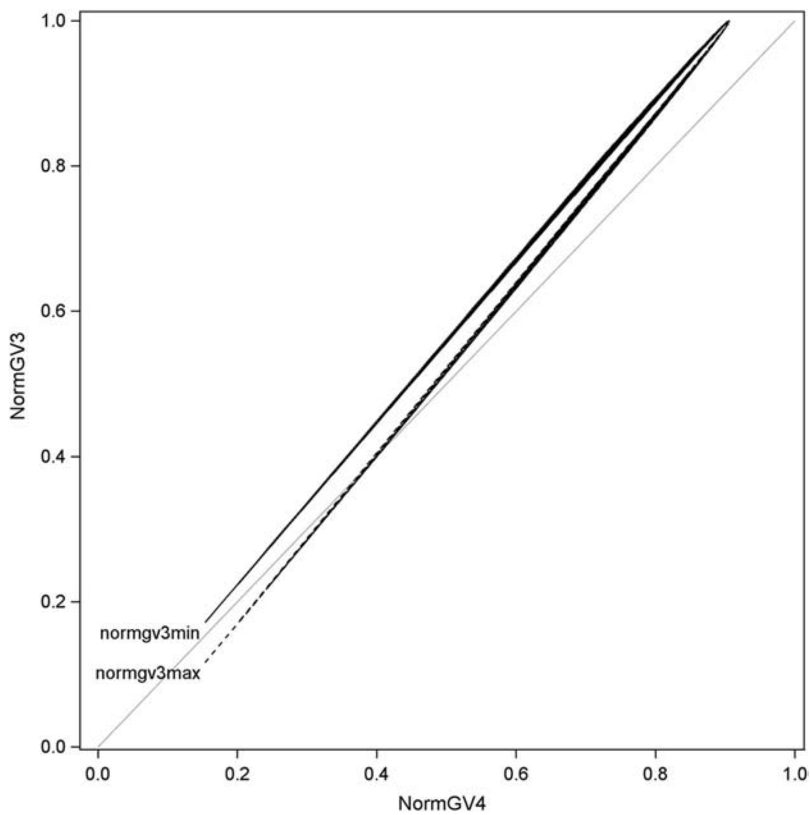

Figure 3. Relative NGV(3 Categories) as a function of relative NGV(4 Categories) for p_1 = .02. NGV = normalized general variance; min = minimum; max = maximum.

religion, marital status, occupation, and political party affiliation. Technically, we looked for a meaningful, and interpretable, measure of scatter for categorical variables, a topic that is surprisingly overlooked in most of our textbooks (see Kader & Perry, 2007), and that has been largely ignored by quantitative psychologists.

Our article demonstrates the importance of choosing an appropriate measure of diversity. Traditionally, investigators have relied on a simple distinction between the majority and the minority groups (all pooled into one). We have argued in the introduction that this simplistic approach is insufficient, and our examples illustrate this point vividly and, we hope, convincingly. We focused on a more complex measure that represents all relevant categories—the GV—whose calculation is straightforward, generalizes in a natural way the well-understood measure of variance of the binomial distribution, has a simple and intuitive interpretation (the likelihood of randomly picking out two individuals from the different groups in the population), and is directly related to the Pearson chi-square test of uniformity and its associated measure of effect size (Johnston et al., 2006). We also described a lesser used, but equally effective, approach to measuring diversity—the entropy of the distribution—that is based on information theory and is related to the G^2 test of uniformity and its associated measure of effect size (Johnston et al., 2006).

GV and entropy are very highly correlated, but each has its own unique and special features, and their dissimilarities may have important implications for one's results, as our examples demonstrate. In the first example, when we normalized both the GV and entropy such that their values share the same end points where 0 stands for homogeneity (lack of diversity) and 1 for maximal heterogeneity (equal proportions in all groups), GV yielded slightly higher values than entropy, and its distribution was slightly more variable. In the second example, we found that comparing the GV and entropy value of each school to some meaningful benchmark yielded slightly different results depending on the choice of measure. Specifically, more schools were classified as "less diverse as the city's population" according to GV than the entropy.

These differences depend on the nature of the distribution in the population in unintuitive and unexpected ways. For example, in binomial distributions (C = 2), NGV and NH become closer to each other as (a) the two proportions are more similar to each other (of course, they coincide when the two categories are equal and NGV = NH = 1.0) and (b) when all the population is concentrated in one of the groups (NGV = NH = 0.0). Interestingly, NGV and NH are farthest apart when one of the groups makes up 90% of the population and the other makes up the remaining 10% (NGV = 0.360 and NH = 0.469). In trinomial distributions (C = 3), NGV and NH are most divergent when the population is equally divided between two of the groups (this is the maximal polarization case) and the third group is empty (NGV = 0.750 and NH = 0.639).

Our NAEP examples illustrate the usefulness of diversity as independent variables in a linear regression. The results showed that both measures of diversity (GV and entropy) contributed significantly to prediction of eighth-grade reading and math scores

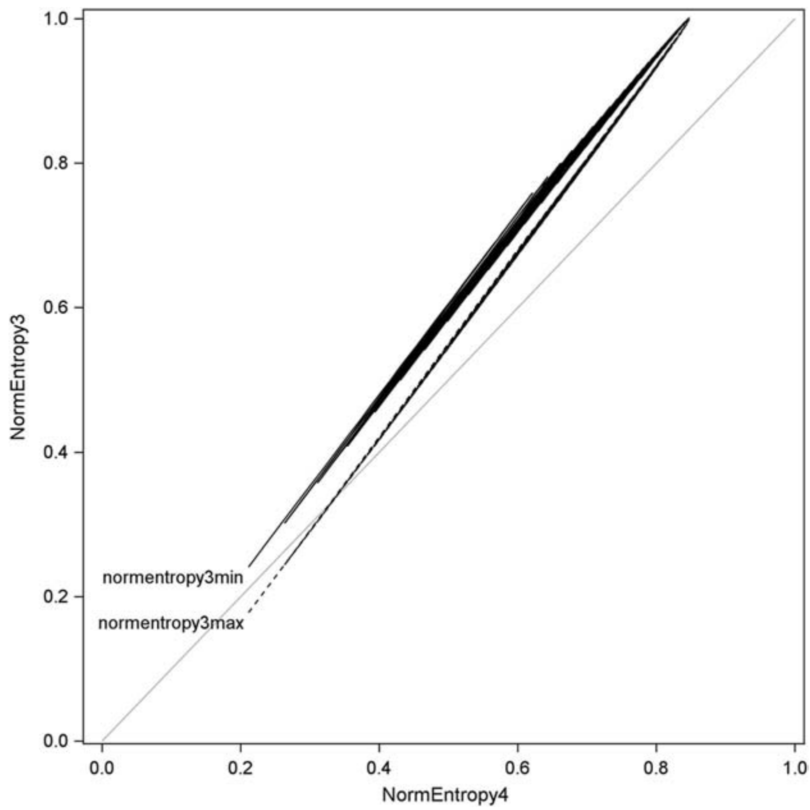

Figure 4. Relative NH(3 Categories) as a function of relative NH(4 Categories) for $p_1 = .02$. NH = normalized entropy; min = minimum; max = maximum.

in the 50 states and the District of Columbia, beyond the proportion of White students, supporting our key claim that the distinction between the majority and the minority is insufficient. Despite the similarity between NGV and NH, the results were not identical: In all cases entropy was a slightly better predictor.

Importantly, in every case researchers must decide how to classify the target population by choosing the number and nature of the categories based on either theoretical or quantitative considerations. In the final examples, we show how the distribution of both the entropy and GV statistics is affected when two or more categories are combined to form a single one, and how data recoding affects the explanatory power of diversity as an independent variable. The key element is to determine whether the extra categories increase the differentiation between the various cases. For example, if we split one of the *C* categories, but the proportion of cases in the new $(C + 1)$ category is, essentially, the same in all groups, neither NGV nor NH would benefit. Furthermore, as the last example shows, in some cases it may be useful to create a separate diversity index for the nonmajority groups. This may be particularly relevant in cases where there are large differences in the proportion of the dominant group across the various units (schools, regions, states, etc.). This last example also illustrates another advantage of the multicategories indices (NGV and NH): their flexibility and versatility that allow one to use them in informative and creative ways in various problems and applications. In fact, it is more appropriate to think about each of them not as a measure of diversity but as a family of measures.

Our intent was to remind researchers of the importance of the appropriate quantification of the concept of diversity; acquaint all researchers in the field with the various possible approaches to its measurement; highlight their properties, strengths, and weaknesses; and illustrate some of the ways in which they can be used to provide answers to important research questions. Our analysis strongly favors the two multicategory measures—GV and entropy—over the simplified majority–minority distinction. These measures are richer, more sensitive, and more informative and, as we have shown in our examples, quite flexible and versatile. Next we address two more subtle questions: Should we use raw measures (GV and H) or their normalized versions (NGV and NH), and should we use GV or entropy?

To Normalize or Not to Normalize?

The normalization step serves only one purpose, namely, to achieve comparability across cases using different number of categories. Thus, in all instances where disparities in the definition of the target attribute are common, it makes sense to normalize. Ethnicity is a prime example, as different organizations and/or different states use, by default, slightly different schemes with various numbers of categories. The price one pays for this generality of the normalized measures is the change in the interpretation. Whereas GV and H are absolute measures of diversity that are expressed in the original units of the target variable (proportions), NGV and NH are relative (or conditional on the number of

Table 5
Summary Measures of the Maximal Change in Relative Diversity as a Function of the Number of Categories and the Size of the Smallest Category

Smallest category	Four categories	Five categories	Six categories
	NGV		
0.02			
N	752	5,952	28,796
Mdn	11.56	6.21	3.89
Central 98%	10.51–12.33	5.55–6.53	3.40–4.07
0.05			
N	574	3,549	12,470
Mdn	10.04	5.32	3.23
Central 98%	7.79–19.95	3.87–11.15	2.23–7.96
	NH		
0.02			
N	752	5,952	28,796
Mdn	18.80	11.57	8.07
Central 98%	15.98–23.06	8.28–14.58	6.98–9.25
0.05			
N	574	3,549	12,470
Mdn	13.59	8.18	5.41
Central 98%	11.58–17.56	6.33–10.22	3.92–6.58

Note. NGV = normalized general variance; NH = normalized entropy.

categories, C) indices. One major appeal of GV is its intuitive and straightforward interpretation—the probability that two randomly selected cases belong to different subgroups. The normalized version, NGV, is a ratio of two probabilities, GV/Max(GV|C) and, as such, is more difficult to interpret in an absolute sense. Therefore, in cases where it is highly likely that all measurements of the target variable be based on (or are easily transformed to) the same number of categories, it is highly recommended not to normalize. Consider, for example, a study of the diversity of marital status in various locations and at various ages. The standard $C = 4$ categories (single/never married, married, divorced, widowed) are defined similarly, and widely accepted, all over the world, and there is no point in normalizing. Of course, this applies to NH as well.

GV or Entropy?

We have shown that the two measures share many properties, and in most cases they are likely to lead to similar conclusions.

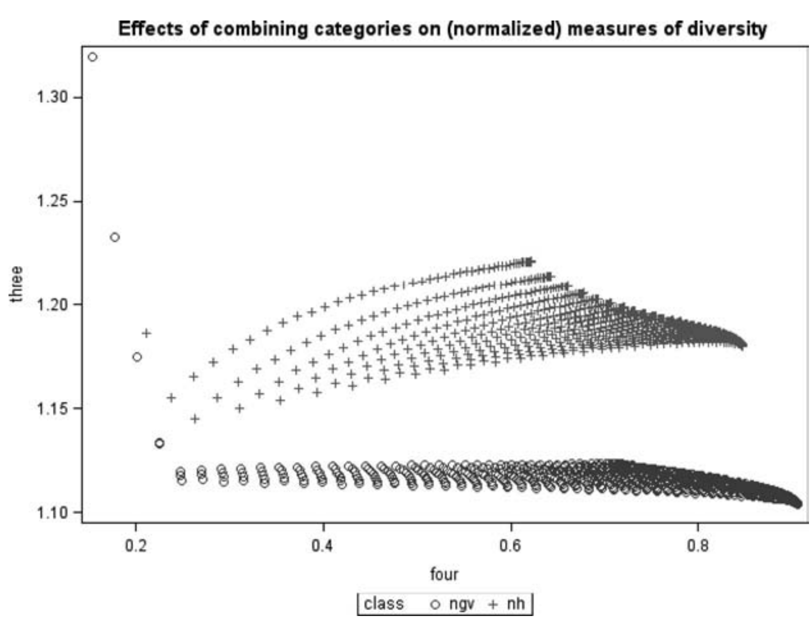

Figure 5. Distribution of the maximal change in relative diversity when the number of categories is reduced from $C = 4$ to $C = 3$ for normalized general variance (NGV) and normalized entropy (NH).

Thus, it is more difficult to justify a principled position in favor of one (and against the other), but practical considerations lead us to favor, slightly, the GV and its normalized version. It is likely that most researchers would find its calculation to be easier, and its probabilistic interpretation to be more intuitive and compelling. Also, given that most researchers are familiar, and comfortable, with the binomial distribution and its variance, GV will be seen by most as a "natural" and direct generalization. Finally, our analysis of robustness of the (normalized) indices suggests that NGV is much less sensitive to the number of categories involved (in particular, see Figure 5).

We hasten to add that the similarities between the two measures are much more prominent than the differences and, in our experience, in most cases they lead to similar results. As we have shown, the choice between them amounts to choosing between Pearson's chi-square and the maximum likelihood G^2, suggesting that there is no universally best solution, and the differences are very small (e.g., Steele, Smart, Hurst, & Chaseling, 2009). Thus, we urge researchers that, whenever possible, they consider both measures. The major difference between the two measures is the scale, and depending on the nature of the particular data being analyzed, one of them may prove to be more useful and informative. Our analysis of the NAEP data where, unexpectedly, the entropy was consistently a better predictor of the scores than the GV is a perfect illustration of this point.

References

Agresti, A., & Agresti, B. F. (1978). Statistical analysis of qualitative variation. In K. F. Schuessler (Ed.), *Sociological methodology* (Vol. 9, pp. 204–237). San Francisco, CA: Jossey-Bass.

Allport, G. W. (1954). *The nature of prejudice.* Reading, MA: Addison-Wesley.

Azen, R., Budescu, D. V., & Reiser, B. (2001). Criticality of predictors in multiple regression. *British Journal of Mathematical and Statistical Psychology, 54,* 201–225. doi:10.1348/000711001159483

Bachi, R. (1956). A statistical analysis of the revival of Hebrew in Israel. In R. Bachi (Ed.), *Scripta Hierosolymitana: Vol. 3. Studies in economics and social sciences* (pp. 179–247). Jerusalem, Israel: Magnum Press.

Bernell, S. L., Mijanovich, T., & Weitzman, B. C. (2009). Does the racial composition of the school environment influence children's body mass index? *Journal of Adolescent Health, 45,* 40–46. doi:10.1016/j.jadohealth.2008.11.013

Biemann, T., & Kearney, E. (2010). Size does matter: How varying group sizes in a sample affect the most common measures of group diversity. *Organizational Research Methods, 13,* 582–599. doi:10.1177/1094428109338875

Biswas, A., & Mandal, S. (2010). Descriptive measures for nominal categorical variables. *Statistics and Probability Letters, 80,* 982–989. doi:10.1016/j.spl.2010.02.012

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure.* New York, NY: Free Press.

Cook, A., & Glass, C. (2008). The impact of LGBT policies on ethnic/racial and gender diversity among business school faculty. *Journal of Diversity in Higher Education, 1,* 193–199. doi:10.1037/a0012987

Dougherty, K. D., & Huyser, K. R. (2008). Racially diverse congregations: Organizational identity and the accommodations of differences. *Journal for the Scientific Study of Religion, 47,* 23–43. doi:10.1111/j.1468-5906.2008.00390.x

Gini, C. (1912). *Variabilità e mutabilità: Contributo alla studio delle distribuzioni e delle relazioni statistiche* [Variability and changeability: Contribution to the study of distributions and statistical relations]. Bologna, Italy: Cuppini.

Gottfredson, N. C., Panter, A. T., Daye, C. E., Allen, W. A., Wightman, L. F., & Deo, M. E. (2008). Does diversity at undergraduate institutions influence student outcomes? *Journal of Diversity in Higher Education, 1,* 80–94. doi:10.1037/1938-8926.1.2.80

Gurin, P. Y., Dey, E. L., Gurin, G., & Hurtado, S. (2003). How does racial/ethnic diversity promote education? *Western Journal of Black Studies, 27,* 20–29.

Gurin, P., Dey, E. L., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review, 72,* 330–367.

Haberman, S. J. (1982). Analysis of dispersion of multinomial responses. *Journal of the American Statistical Association, 77,* 568–580. doi:10.2307/2287713

Hallinan, M. T., & Teixeira, R. A. (1987). Opportunities and constraints: Black–White differences in the formation of interracial friendships. *Child Development, 58,* 1358–1371. doi:10.2307/1130627

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review, 32,* 1199–1228. doi:10.5465/AMR.2007.26586096

Hirshman, A. O. (1964). The paternity of an index. *American Economic Review, 54,* 761.

Hunt, M. O., Wise, L. A., Jipguep, M.-C., Cozier, Y. C., & Rosenberg, L. (2007). Neighborhood racial composition and perceptions of racial discrimination: Evidence from the Black women's health study. *Social Psychology Quarterly, 70,* 272–289. doi:10.1177/019027250707000306

Hurtado, S. (2005). The next generation of diversity and intergroup relations research. *Journal of Social Issues, 61,* 595–610. doi:10.1111/j.1540-4560.2005.00422.x

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Johnston, J. E., Berry, K. J., & Mielke, P. W., Jr. (2006). Measures of effect size for the chi-squared and likelihood-ratio goodness-of-fit tests. *Perceptual and Motor Skills, 103,* 412–414. doi:10.2466/pms.103.2.412-414

Juvonen, J., Nishina, A., & Graham, S. (2006). Ethnic diversity and perceptions of safety in urban middle schools. *Psychological Science, 17,* 393–400. doi:10.1111/j.1467-9280.2006.01718.x

Kader, G. D., & Perry, M. (2007). Variability for categorical variables. *Journal of Statistics Education, 15*(2). Retrieved from http://www.amstat.org/publications/jse/v15n2/kader.html

Lieberson, S. (1969). Measuring population diversity. *American Sociological Review, 34,* 850–862. doi:10.2307/2095977

Miethe, T. D., & McDowall, D. (1993). Contextual effects in models of criminal victimization. *Social Forces, 71,* 741–759. doi:10.2307/2579893

Noble, E. E., & Sanchez, P. P. (1993). A note on the information content of a consistent pairwise comparisons judgment matrix of an AHP decision maker. *Theory and Decision, 34,* 99–108. doi:10.1007/BF01074896

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90,* 751–783. doi:10.1037/0022-3514.90.5.751

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures* (A. Rosin, Trans.). Oxford, England: Viking.

Rice, K. J., & Csmith, W. R. (2002). Sociological models of automotive theft: Integrating routine activity and social disorganization approaches. *Journal of Research in Crime and Delinquency, 39,* 304–336. doi:10.1177/002242780203900303

Richard, O. C., Murthi, B. P. S., & Ismail, K. (2007). The impact of racial diversity on intermediate and long-term performance: The moderating role of environmental context. *Strategic Management Journal, 28,* 1213–1233. doi:10.1002/smj.633

Seaton, E. K., & Yip, T. (2009). School and neighborhood contexts, perceptions of racial discrimination, and psychological well-being among African American adolescents. *Journal of Youth and Adolescence, 38,* 153–163. doi:10.1007/s10964-008-9356-x

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27,* 379–423, 623–656.

Simpson, E. H. (1949). Measurement of diversity. *Nature, 163,* 688. doi:10.1038/163688a0

Sommers, S. R., Warp, L. S., & Mahoney, C. C. (2008). Cognitive effects of racial diversity: White individuals' information processing in heterogeneous groups. *Journal of Experimental Social Psychology, 44,* 1129–1136. doi:10.1016/j.jesp.2008.01.003

Steele, M., Smart, N., Hurst, C., & Chaseling. J. (2009). Evaluating the statistical power of goodness-of-fit tests for health and medicine survey data. In R. S. Anderssen, R. D. Braddock, & L. T. H. Newham (Eds.), *18th IMACS World Congress and MODSIM09 International Congress on Modelling and Simulation* (pp. 192–196). Canberra, Australia: Modelling and Simulation Society of Australia and New Zealand.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33,* 1–39. doi:10.1146/annurev.ps.33.020182.000245

Teachman, J. D. (1990). Analysis of population diversity: Measures of qualitative variation. *Sociological Methods and Research, 8,* 341–362. doi:10.1177/004912418000800305

Walsh, J. A., & Taylor, R. B. (2007). Predicting decade-long changes in community motor vehicle theft rates: Impact of structure and surround. *Journal of Research in Crime and Delinquency, 44,* 64–90. doi:10.1177/0022427806295552

Welch, S., Sigelman, L., Bledsoe, T., & Combs, M. (2001). *Race and place: Race relations in an American city.* New York, NY: Cambridge University Press.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale NJ: Erlbaum.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9,* 1–27. doi:10.1037/h0025848

Received June 12, 2010
Revision received June 15, 2011
Accepted November 28, 2011 ■