California Center for Population Research
University of California - Los Angeles

Identifying Organizational Faultlines with Latent Class Cluster Analysis

Barbara S. Lawrence
Michael J. Zyphur

CCPR-2009-030

December 2009
Last Revised: September 2009

California Center for Population Research
On-Line Working Paper Series

IDENTIFYING ORGANIZATIONAL FAULTLINES

WITH LATENT CLASS CLUSTER ANALYSIS

Barbara S. Lawrence
Anderson Graduate School of Management
University of California Los Angeles
Los Angeles, CA 90095-1481
barbara.lawrence@anderson.ucla.edu

Michael J. Zyphur
University of Melbourne
Department of Management and Marketing
Parkville 3010
AUSTRALIA
mzyphur@unimelb.edu.aiu

September 10, 2009

We thank the organization in which these data were collected for its cooperation and support of this research. Major funding for computational analysis was provided by the National University of Singapore Business School, and Singapore's Ministry of Education to the second author. The authors would like to thank these institutions for their support.

ABSTRACT

Faultline theory proposes that when multiple attributes are aligned in groups they create homogeneous subgroups, characterized by within-group similarities and between-group differences. As homogeneity increases, these differences are increasingly likely to acquire meaning to subgroup members and thus to influence behavior. While the face validity of faultlines is theoretically appealing, empirical measures have been difficult. The most commonly used, *Fau, D,* and *FLS*, have been limited to small groups, two or at most three attributes, and do not easily integrate nominal, categorical, and continuous variables. This paper proposes latent class cluster analysis (LCCA) as an additional analytical tool. LCCA is useful for large groups, and facilitates analysis of numerous attributes independent of underlying distributions. After reviewing the multiple-attribute literature, the most common faultline measures are described and compared with LCCA. A study of faultlines in a large organization is presented. LCCA induces a five-class model of organizational faultlines. A comparison of work-related communication contacts indicates that subjects have more within-subgroup than between-subgroup contacts, supporting the criterion-related validity of the faultline solution.

Scholars have long been interested in the distribution of individuals' demographic attributes in social systems. These distributions create a distinctive social context to which individuals respond. People with similar attributes, such as gender or age, tend to recognize themselves as distinct from others, often creating psychologically salient and socially meaningful groups that influence work. Research connects the demographic distributions characterizing such groups to a long list of individual, group and organizational outcomes including conflict (Pelled, Eisenhardt, & Xin, 1999), turnover (Elvira & Cohen, 2001), performance (Cannella, Park, & Lee, 2008), corporate foreign investment (Barkema & Shvyrkov, 2007), career mobility (L. E. Cohen, Broschak, & Haveman, 1998) and salary (Castilla, 2008). Although many studies focus on the distribution of one demographic attribute, others, such as faultline research, involve more.

Faultline theory (Lau & Murnighan, 1998) proposes that the distribution of demographic attributes creates subgroups in small social systems, such as work groups. The more closely attributes are aligned, the stronger the faultline. These alignments produce a structural constraint on relationships that maximizes similarities within and differences between subgroups by priming the aligned attributes, increasing their salience for group members. As a result, faultlines represent the boundaries separating behaviorally-meaningful subgroups. Li and Hambrick (2005), for instance, examined factional faultlines in the top management teams of 71 international joint ventures. They found that increasing faultline size, as measured by the age, tenure, gender and ethnic differences between expatriate and local managers in each team, was related to increasing emotional and task conflict.

The distinguishing feature of faultlines is the assumption that individuals' attributes acquire meaning interdependently. In other words, faultlines are identified by the joint distribution of several attributes rather than the individual distributions of each. This interdependence distinguishes faultline theory from many multiple attribute approaches, which assume independence and measure additive or averaged distributional effects (e.g., Urada, Stenstrom, & Miller, 2007). Such independent effects are relatively easy to study with standard techniques such as regression and analysis of variance. However, interdependence presents measurement issues that have limited the development of faultline research and theory.

This study examines latent class cluster analysis (LCCA) as a method for identifying an organization's faultlines, alignments of individual attributes that define relatively large, meaningful subroups conditioned by how individuals perceive their organization. LCCA provides an alternative to existing methods. It permits inclusion of many attributes at a time and facilitates study of nominal, categorical and continuous attributes in the same analysis, a difficulty noted in previous work (Lau & Murnighan, 1998). In addition, it is suited for analyzing large social systems, such as organizations. The discussion below is organized as follows. We begin with a review of existing multiple attribute theories that assume interdependence, focusing on their similarities and differences in identifying attribute-based subgroups. This is followed by a longer evaluation of existing faultline measures, focusing on identification rather than on strength or distance. We conclude with an empirical study of a large organization. We use LCCA to identify organizational faultlines and then provide criterion-related validity using work-related communication contacts within and between the subgroups these faultlines define.

Multiple Attribute Interdependence

Support for faultline theory appears across several disciplines, with some studies emphasizing psychological mechanisms and others focusing on structural explanations. Distinctiveness and crossed-category theory emphasize psychological mechanisms as producing multiple attribute effects. Distinctiveness theory (McGuire, McGuire, & Winton, 1979) suggests that when presented with a substantial quantity of complex information, individuals selectively perceive attributes that appear distinctive within the social context. In an organizational context, when an individual belongs to two minority groups, he or she will identify more strongly with the smaller of the two groups. Mehra et. al (1998) examined a second-year class of MBA students in which white men represent 60%, white women represent 28%, and minority men and women represent 6% each of the study population (sample N=159). They asked students to indicate with whom they identified and who were their friends. Whites were more likely to identify and be friends with others on the basis of sex rather than race; for Whites, sex is a more distinctive attribute than race. In contrast, minorities were more likely to identify and be friends with others on the basis of race rather than sex; for minorities, race is the more distinctive category. In each case, students connected with others using the most, rather than least, distinctive of their own attributes. This differs from an additive approach because it suggests that individuals alter their behavior with others depending on the distribution of those attributes in their social context.

Crossed-attribute categorization presents a related approach to multiple attribute interdependencies. When individuals share more than one attribute and when no one attribute is dominant, crossed-attribute categories emerge (Ashforth & Johnson, 2001; Vescio, Hewstone, Crisp, & Rubin, 1999) such as gender-age (Klauer, Ehrenberg, & Wegener, 2003). Although this literature typically involves dyads rather than groups, the results are suggestive. Urada et al. (2007) propose a feature detection strategy, consistent with distinctiveness theory, suggesting that people use similarity thresholds for evaluation. When individuals perceive a target as enough-like-me, the target gets defined as an ingroup member, independent of the number of attributes involved. This contrasts with an additive approach, in which each homophilous attribute increases individuals' similarity perceptions. The feature detection strategy is particularly relevant to organizational decisions that relate to performance, such as promotions and salary. Managers evaluate employees in a salient social context in which correlated attributes acquire significant meaning because they are related to performance.

Intersectionality and consolidation theory emphasize structural explanations as the mechanism producing multiple attribute effects. These explanations suggest that multiple attribute effects result because the distribution of attributes in a population both constrains and facilitates individuals' opportunities to become aware of and develop relationships with one another. This does not exclude psychological mechanisms, but it emphasizes that these mechanisms are strongly influenced by the relationships among the demographic attributes that define social context. For instance, Black and feminist sociologists use intersectionality to explain the joint effects of gender and race (Browne & Misra, 2003). In this work, intersecting categories are socially constructed through historical or local social contexts: "Race is 'gendered' and gender is 'racialized,' so that race and gender fuse to create unique experiences and opportunities for all groups—not just women of color" (2003, p. 488). These experiences are shaped by ideology, control of economic and political resources and the unequal distribution of valued resources across subgroups.

This approach is consistent with Blau's concept of consolidation (1977), the strength of the positive relationship among attributes (p. 276). Blau (1977) suggests that a social system's heterogeneity and inequality are defined by the distribution of individuals' multiple attributes. As consolidation increases, the number of groups comprising the social structure decreases, and this decreases intergroup social interactions. Decreased intergroup interactions produce both fewer cordial and fewer conflictual associations. Thus, structural opportunity represents the basic mechanism by which multiple attributes influence behavior. Consolidated nominal attributes, such as gender or ethnicity, produce lower heterogeneity; consolidated graduated attributes, such as age and tenure, produce higher status differences.

Faultline theory suggests a mixture of social psychological and structural mechanisms. Lau & Murnighan (1998) define faultlines as boundaries or break points identified by the alignment of one or more individual attributes that separate a group into distinct subgroups that hold meaning for their members. This approach emphasizes the structural opportunity created by intersecting multiple attributes as well as the psychological identities that facilitate meaningful subgroups. When people identify themselves by attributes such as age, race, and gender, they are likely to psychologically-orient themselves towards others who share those attributes. Attributes that are both salient and apparent to group members are likely to drive subgroup formation (Lau & Murnighan, 1998, p. 328). As similarities within and differences between clusters of individuals are found along more and more attributes, the potential for intra-cluster alignment and inter-cluster difference increases. This literature is at an early stage of development and appropriate measures are still emerging. Our focus is on identifying faultlines in large social systems and assessing the validity of the subgroups they create.

Organizational Faultlines vs. Workgroup Faultlines

In this study, we examine whether LCCA identifies sets of aligned attributes that define meaningfully-distinct subgroups of individuals in a large organization. This analysis differs from earlier faultline work because the focus is on large social systems rather than on small groups. The theoretical rationale for organizational faultlines and their effects remains similar to that associated with extant faultline research: aligned attributes produce subgroups that are socially-meaningful to subgroup members and this produces distinct within-subgroup and between-subgroup behavior. However, there are several differences that complicate and perhaps prohibit organizational faultline analysis using existing methods.

One is that members of small groups are typically known, whereas the boundaries of larger, informal social structures are less clear. In small groups, everyone knows everyone else, and while members may not know their subgroup boundaries, they do know which individuals define the group (although see Mortensen, 2008). In large organizations, individuals do not know everyone and the people they do know are unlikely to represent a random sample. These non-random others constitute an individual's organizational reference group, the apparent "organization" as he or she perceives it (Lawrence, 2006). Organizational reference groups include all the others of whom an individual is aware, including close, distant, and no associations. This awareness criterion suggests that an individual's organizational reference group represents a portrait of the organization as he or she observes and experiences it. Thus, if we identify faultlines in the collective set of individuals' organizational reference groups then subgroup membership is likely to be meaningful.

Another difference involves the size of the subgroups created by faultlines. Organizational faultlines identified in this collective of organizational reference groups should define discrete subgroups similar to those in the existing literature. However, in contrast to that literature where subgroups represent small segments of already small groups, organizational faultlines are likely to create large, discrete subgroups, which exceed the size of small groups and perhaps even of small organizations. As a result, the mechanisms that operate within and between these subgroups provide a picture of a larger, informal social structure than the subgroups identified within work groups.

Extant Methods for Identifying Faultlines

A central characteristic of faultlines is that they are latent, unobserved or informal boundaries, which by definition define latent, unobserved or informal subgroups. Lau and Murnighan (1998) suggest that faultline strength increases with the increasing homogeneity of the subgroups they create—thereby increasing the probability that subgroup membership will influence individual behavior. Such homogeneity is conceptualized using both the similarity among individuals' attributes within a given subgroup and their collective differences from those in another subgroup. Thus, the homogeneity associated with faultline identification is indexed by examining the proportion of between-subgroups variance to within-subgroups variance.

The current faultline literature includes several approaches to measuring homogeneity. The majority of these studies involve experimental methods, in which faultlines and subgroup boundaries are defined a priori (cf. Homan, et al., 2008; Homan, van Knippenberg, Van Kleef, & De Dreu, 2007; Lau & Murnighan, 2005; Pearsall, Ellis, & Evans, 2008; Polzer, Crisp, Jarvenpaa, & Kim, 2006; Rico, Molleman, Sanchez-Manzanares, & Van der Vegt, 2007; Sawyer, Houlette, & Yeagley, 2006). However, several faultline measures have been proposed for situations where subgroup boundaries are unknown. Although their computational capabilities are still being developed, at present, each measure has several empirical limitations (See Table 1). They have been restricted to small social systems, such as work groups. They accommodate analyses of small numbers of attributes, generally two or three, because the methods become substantially more intractable as the number of attributes increases. Further, when mixing nominal, categorical and continuous variables, they require categorizing or weighting the variables in terms of their relative importance in faultline formation, which requires a number of assumptions. LCCA addresses many of these issues, suggesting an alternate method for identifying faultlines.

Table 1 About Here

Lau and Murnighan (1998)

Although Lau & Murnighan (1998) do not provide a method for identifying faultlines, they do discuss the analytical issues. They suggest that "measures of demographic diversity within a group must be dispersion indexes" (1998, p. 327), such as a modification of Blau's measure of diversity or others based on Euclidean distances across people. They also state that such measures should not combine nominal, categorical, and continuous measures because it "would be like cross-fertilizing apples and oranges" (1998, p. 327). This separation of categorical and continuous measures likely represents empirical limitations of existing dispersion indices rather than theoretical exclusion.

Thatcher, Jehn, and Zanutto (2003)

In the first work to develop a method for determining faultlines, Thatcher, Jehn, and Zanutto (2003) proposed *Fau*, which is a measure of the "percent of total variation in overall group characteristics accounted for by the strongest group split" (p. 225). Essentially, *Fau* is the proportion of between-subgroups variance to total variance, and can be shown as

$$Fau_g = \left(\frac{\sum_{j=1}^{p} \sum_{k=1}^{2} n^g{}_k \, (y_{\cdot jk} - y_{\cdot j \cdot})^2}{\sum_{j=1}^{p} \sum_{k=1}^{2} \sum_{i=1}^{n^g{}_k} (y_{ijk} - y_{\cdot j \cdot})^2} \right) \qquad g = 1, 2, \ldots S \qquad (1)$$

where y_{ijk} is the value along the *j*th attribute for the *i*th member in the *k*th subgroup, $y_{\cdot j \cdot}$ is the mean of the *j*th attribute, $y_{\cdot jk}$ is the mean of the *j*th attribute in the *k*th subgroup, n is the number of people in a known group, p is the number of attributes along which the group members have been measured, and n^g_k is the number of people in the *k*th subgroup of the *g*th split into subgroups. *Fau* is then calculated for *S* many splits *g*, where

$$S = 2^{n-1} - 1 \quad (2)$$

Following the computation of *Fau* for all the *g* splits, Thatcher et al. (2003) recommend choosing the group split that produces the largest overall proportion of between-subgroups to total variance. This split, in which Fau_g is closest to 1.0, represents the faultline and defines the two subgroups. *Fau* thus reflects faultline theory by incorporating the ideas that faultline strength and its impact on behavior increases with the increasing homogeneity of subgroups.

Thatcher et al. (2003) recommend using *Fau* to identify one faultline and two subgroups when examining known groups (see Lau & Murnighan, 2005 for application in experimental design). They acknowledge that Lau and Murnighan's (1998) original conceptualization of faultlines allows for more than one faultline and more than two subgroups. However, they (2003) suggest identifying only one faultline for two reasons. First, groups frequently include only a few individuals, making more than two meaningful subgroups unlikely, and second, the computational complexity of *Fau* increases greatly with more than one faultline. Authors using this work have maintained this perspective (e.g., Molleman, 2005).

Somewhat in opposition to Lau and Murnighan's (1998) recommendations, Thatcher et al. (2003) propose a *Fau* scaling scheme that allows the simultaneous use of continuous and non-continuous attributes. This scheme is designed to assign equal weights to differences across people along continuous and non-continuous variables. To accomplish this, Thatcher et al. (2003) recommend turning any non-continuous variables with *c* categories into *c* variables that are dummy coded (0 or $1/\sqrt{2}$) to express individuals' standing along the variable of interest. The exception occurs when $c = 2$, in which case there is only a single dummy-coded variable.

Dividing the usual dummy-code of 1 by √2 allows a difference between two people along the categorical variable to count as 1 unit of difference when summing across all attributes. Continuous variables are divided by a value that is specific to a given attribute and based on a researcher's theory regarding the importance of a given distance along the variable. For example, if a researcher thought that 10 years of difference in age was equivalent to a difference in gender, then the researcher would divide all individuals' ages by 10. This would allow a difference of 10 years in age to have the same weight in *Fau* as any difference along non-continuous variables. Thatcher et al. (2003) acknowledge the large amount of subjectivity in this process, but see it as a necessary requirement for discerning faultlines. Although Fau has been limited to small groups, a recent modification may facilitate its use in larger populations, such as organizations (K. Bezrukova, personal communication, August 11, 2009).

Shaw (2004)

Another method for identifying faultlines was outlined by Shaw (2004). Shaw bases his measure on the idea that people perceive continuous attributes in meaningfully discrete categories. For example, in some contexts anyone between the ages of 20 and 35 may be considered "young." As he notes, "a substantial body of literature suggests that cognitive categorization processes naturally occur when individuals of different characteristics interact in groups" (2004, p. 70). Researchers accomplish this either by using extant research to deduce meaningful categories or by using an empirical approach to induce them, for example by asking participants about the cognitive categories they use to classify others.

Faultlines are then assessed by examining the "internal alignment" (IA) of a given variable, defined as "the extent to which members within a particular subgroup are similar to one another on all other relevant variables" (Shaw, 2004, p. 72). This is done by comparing observed attribute distributions to expected null distributions across the categories of other variables where

$$\mathrm{IA}_{A1/B/\mathrm{OBS}} = \sum_{c=1}^{m}(\mathrm{O}_{A1c} - \mathrm{E}_{A1c})^2/\mathrm{E}_{A1c} \qquad (3)$$

$\mathrm{IA}_{A1/B/\mathrm{OBS}}$ is the observed internal alignment for a category *1* (e.g., males) for a variable *A* (e.g., gender) across all of the *m* categories of a variable *B* (e.g., Caucasian and African-American for a variable Race), O_{A1c} is the observed number of people in a category *1* (e.g., males) for a variable *A* (e.g., gender) in the *c*th category (e.g., caucasion) for a variable *B* (e.g., race), and E_{A1c} is the expected number of people in a category *1* for a variable *A* in the *c*th category for a variable *B*—with the assumption that such an expectation takes the form of a random distribution. By summing across all *m* categories, that is by summing across each *c*, and after taking the square of the distance between O and E, one has computed the sum of the squared differences between the observed and the expected data. Dividing this sum by E allows $\mathrm{IA}_{A1/B/\mathrm{OBS}}$ to equal the ratio of the squared differences between observation and expectation to expectation and is analogous to the computation of a χ^2 statistic.

This means that the ratio describes the distribution of individuals within category *1* (e.g., males) of variable *A* (e.g., race) across all *m* categories of variable *B* (e.g., across all of the race categories). The ratio moves towards 0.0 as individuals within category *1* of variable *A* tend to be randomly distributed across the categories of variable *B*. The ratio increases above 0.0, heading towards "perfect alignment," as people within category *1* of variable *A* become more systematically distributed across the *m* categories of variable *B*.

However, $\mathrm{IA}_{A1/B/\mathrm{OBS}}$ is not yet useful for faultline assessment because $\mathrm{IA}_{A1/B/\mathrm{OBS}}$ only indexes a ratio of squared deviations and does not take into account the maximum and minimum internal alignments that are possible. It is important to do this because these values change with different numbers of people in a given group and with different numbers of categories for any variables of interest. These issues are solved through the following

$$IA_{A1/B} = (IA_{A1/B/\mathrm{OBS}} - IA_{A1/B/NONALLIGN})/\mathrm{MaxDiff} \quad (4)$$

where

$$\mathrm{MaxDiff} = IA_{A1/B/PERFECT} - IA_{A1/B/NONALLIGN} \quad (5)$$

and $IA_{A1/B/NONALLIGN}$ is the internal alignment score when there is perfect *non*alignment for individuals within a category *1* along a variable *A* across the *m* categories of a variable *B* $IA_{A1/B/PERFECT}$ is the internal alignment score when there is perfect alignment for individuals within a category *1* along a variable *A* across the *m* categories of a variable *B*, and, thus, MaxDiff represents the possible range of internal alignment scores. Therefore, $IA_{A1/B}$ is the ratio of observed internal alignment to the possible internal alignment score.

To derive the average alignment score across all of the categories of a variable *A*, researchers may then compute the average internal alignment score, $IA_{A/B}$, across all categories of variable *A*. In an example with three categories of the variable *A*

$$IA_{A/B} = (IA_{A1/B} + IA_{A2/B} + IA_{A3/B})/3 \quad (6)$$

This may further collapse across multiple other variables beyond *B* with, for example,

$$IA_{A} = (IA_{A/B} + IA_{A/C} + IA_{A/D})/3 \quad (7)$$

and may integrate across all variables with, for example,

$$IA_{OVERALL} = (IA_{A} + IA_{B} + IA_{C})/3 \quad (8)$$

Beyond this intuitive understanding of internal alignment scores, it is important to consider that $IA_{A/B}$, IA_{A}, and $IA_{OVERALL}$ are not the only values required to understand faultlines. This results because IA references only the degree to which individuals within a given category, or multiple categories, are aligned across other categories. It does not provide insight into the degree to which individuals in other categories have similar cross-category memberships. This is important because, according to Lau and Murnighan (1998), to the extent that the people *outside* of category *1* of a variable *A* share the same membership along a variable *B* as the people *inside* of category *1* of a variable *B*, the IA score of the people inside category *1* of variable *A* will be less meaningful. Therefore, it is important to consider the "cross-subgroup alignment index" (CGAI), which is defined as the degree of similar category memberships (e.g., being in the *c*th category of a variable *B*) for people in different subgroups (e.g., people in category *1* versus *2* for a variable *A*), where CGAI ranges between 0.0 and 1.0, that is, no cross-subgroup alignments versus perfect cross-subgroup alignments, respectively. When CGAI is low it means that the IA score of a subgroup is more meaningful than when CGAI is high. Shaw (2004) recommends weighting IA as follows

$$FLS = \mathrm{IA} \cdot (1 - \mathrm{CGAI}) \tag{9}$$

with a more specific formula being expressed as

$$FLS_A = \mathrm{IA}_A \cdot (1 - \mathrm{CGAI}_A) \tag{10}$$

and with an overall formula

$$FLS_{OVERALL} = \mathrm{IA}_{OVERALL} \cdot (1 - \mathrm{CGAI}_{OVERALL}) \tag{11}$$

where all terms may be understood as previously mentioned.

In sum, Shaw (2004) outlines a method for the computation of values of the internal alignment of attributes within subgroups and a method for their aggregation and weighting.

Li and Hambrick (2005)

A recent article by Li and Hambrick (2005) explored faultlines in 71 top management teams participating in joint ventures. The teams they studied each involved two known, factional subgroups: one consisting of local managers and the other of expatriates. Starting with these known factions, Li and Hambrick (2005) outline a computational logic to compare the further differences of these subgroups along four attributes: age, gender, tenure and ethnicity. This logic computes the faultline size between the two factional subgroups, which they describe as having a large value when "two factions differ in their averages [along a given attribute] *and* each faction is tightly clustered around its own average" (p. 804). This definition is equivalent to the criterion of subgroup homogeneity as a precondition for faultline identification, meaning that Li and Hambrick's (2005) faultline size is similar to Lau and Murnighan (1998)'s faultline strength.

To measure faultline size, Li and Hambrick (2005) modified the well known *d* statistic (see J. Cohen, 1988) as a "demographic difference" *d* with the following formula

$$d_1 = \frac{|X_A - X_B|}{\frac{\sigma_A \, \sigma_B}{2} + 1} \qquad (12)$$

where d_1 is the demographic difference between two subgroups along a variable *1*, X_A is the mean along a variable for a subgroup *A*, X_B is the mean along a variable for a subgroup *B*, σ_A is the standard deviation of a variable *A*, σ_B is the standard deviation of a variable *B*, and a constant is added to the denominator to assure that *d* is a real number when σ_A and σ_B are both equal to zero. This formula allows for the computation of *d* for either continuous or dichotomous variables. Further, by combining across multiple variables researchers may compute an overall faultline score for a given group. However, to do so, Li and Hambrick (2005) recommend first standardizing the *d*s for each variable, such that each group's *d* for a given variable is a standardized deviation away from the average *d* across all groups for that variable.

Bezrukova, Jehn, and Zanutto (2009)

A final technique is presented by Bezrukova, Jehn, and Zanutto (2009). These authors explore the difference between faultline strength and faultline distance, suggesting that faultlines are multidimensional. They note that faultline strength captures the alignment of demographic or other attributes within a group, whereas faultline distance references the magnitude of the difference between subgroups along the attributes of interest.

To measure faultline strength, they use *Fau*, because it is a measure of the ratio of between-subgroups to total variance. As a result, *Fau* captures the homogeneity of subgroups (i.e., the cleanness of their split). To measure faultline distance between subgroups identified with *Fau*, they suggest taking the Euclidean distance between the centroids of the two subgroups' multivariate distributions for the attributes in question. In other words, they examine the distance between the vector of means for the variables that have been used in the computation of *Fau* (Molleman, 2005). This can be shown as

$$D_g = \sqrt{\sum_{j=1}^{p} (y_{1j\cdot} - y_{2j\cdot})^2} \tag{13}$$

where D_g is the multivariate distance between the two subgroups, that is, the difference in their centroids, y_{1j} is the mean for a subgroup *1* along the *j*th attribute, and y_{1j} is the mean for a subgroup *2* along the *j*th attribute. To find D_g requires summing across all of the *p* attributes.

By taking *D* and *Fau* separately, Bezrukova et al. (2009) account for the possible multidimensional nature of group faultlines. *Fau* describes the proportion of the variance between the subgroups in relation to the total variation, which is insensitive to the actual distance between the subgroups on the measures of interest. *D* describes the actual distance between the subgroups along the attributes of interest, which is insensitive to the proportion of variance this distance accounts for in relation to the total variance. Forming the interaction between *Fau* and *D* allows modeling the joint effect of these two aspects of a group's faultline.

Similarities and Differences Across Faultline Measures

These measures of faultlines have both similarities and differences. All of them, except *d*, begin with the assumption that true faultlines within a known group are unknown or latent. Two of them, *Fau* and *FLS,* make the simplifying assumption that each group has only one or at most a few faultlines. For instance, Thatcher et al. (2003, p. 447) propose there is likely one meaningful faultline in any known group. As a result, they recommend using the faultline that allows the greatest ratio of between-subgroups variance to total variance across multiple attributes for two subgroups. Alternately, Shaw's (2004) method requires the formation of subgroups along each attribute based on categorical differences across the attributes of interest.

By adding D_g to measure faultlines in conjunction with *Fau*, Bezrukova et al. (2009) acknowledge that *Fau* only measures proportions of variance, and so they recommend the incorporation of a measure of the actual distance between the subgroups along the variables of interest with D_g. This is also inherent in Shaw's (2004) *FLS*, but in a different manner. Shaw captures the degree to which individuals in a subgroup are aligned across multiple variables with his IA score. This captures within-subgroups variance, but weights the information about variables with the degree to which individuals within other subgroups are misaligned with members of that subgroup. This weighting is done with CGAI, which can be thought of as capturing between-subgroups variance. While Shaw's method differs from *Fau* and D_g, the logic of attempting to discern "the number of demographic attributes that group members align on" and "how far apart these aligned groups are from each other" (Bezrukova et al., p. 5) is also embedded within *FLS*.

Another similarity among these methods is that none explicitly specify a model that best fits the observed data. *Fau* identifies a single faultline using the amount of between-group variance it creates. The fit of any model to the original data is unknown outside an R^2 statistic, which is similar to a fit assessment but limited to the one faultline. This assessment becomes increasingly complex when including discontinuous variables because, after transformation, fitting an estimated model to the original data is not possible. *FLS* computes an overall faultline score by examining attributes' internal alignment. This requires categorizing continuous variables into discontinuous variables. *FLS* does not identify which individuals belong in which subgroups and, therefore, does not produce a model that can be compared to observed data.

LCCA and Organizational Faultlines

What Lau and Murnighan (1998) explore as a subgroup created by a faultline has been explored elsewhere as a "latent class" derived from a LCCA (LCCA; DiStefano & Kamphaus, 2006). A latent class is a group of individuals who exhibit more homogeneity as a cluster, along multiple attributes, than the known group from which they are drawn. This homogeneity is not directly observed but inferred. Statistically, this means that within a known group of individuals, there is likely to be linear dependence among their attributes, that is, unobserved heterogeneity. Lau and Murnighan describe this as "collinearity" among traits that are "correlated" (1998, p. 328) such that various people can be clustered together to form meaningfully homogenous subgroups. This fits with Lau and Murnighan's statement that "group faultlines increase in strength as more attributes are highly correlated, reducing the number and increasing the homogeneity of the resulting subgroups." (1998, p. 328). Flache & Mas (2008) provide an example operationalizing these ideas using a computational model.

The LCCA procedure treats participants' probabilities of membership in latent classes as missing. These probabilities are estimated in an iterative process that produces a model with the best fit to the observed data. With an interest towards model parsimony, models with different numbers of classes are compared along both statistical and substantive grounds (see Muthen, 2003) to choose a final latent class structure. Conveniently, LCCA allows for the integration of continuous and categorical variables without sacrificing any information in the variables—a linking function is used with non-continuous variables. This approach to clustering has benefits over more traditional forms, such as k-means cluster analysis, because the results are not adversely affected by the scale and variance of observed variables (see DiStefano & Kamphaus, 2006; Hagennars & McCutcheon, 2002). This is also true when comparing LCCA to the faultline methods discussed above where continuous variables must be made discontinuous or discontinuous variables must be transformed. Therefore, LCCA may be useful for faultline analysis because researchers are neither forced to make assumptions about the importance of variables, as required in *Fau*, nor required to categorize continuous variables, as in *FLS*.

Although, LCCA can require more individuals than there are observed variables in some statistical packages, this is not a requirement with a full information maximum likelihood estimator (Enders, 2001; see similar thought in Hamaker, Dolan, & Molenaar, 2003). Large samples are required when generalizing to a larger population, with additional individuals allowing for more stable class enumeration and unbiased estimates. However, when such generalizations are not desired it is possible to use LCCA in small groups (see Barkema & Shvyrkov, 2007 for an example). As a result, LCCA is an ideal technique for uncovering faultlines in large groups such as organizations and small groups such as those studied in traditional faultline research.

Method

This study illustrates the use of LCCA in identifying organizational faultlines. We validate the subgroups by comparing subjects' work-related communication contacts within- and between-subgroups. If the faultlines are valid, we should observe greater within-subgroup than between-subgroup contacts as this provides an indication that the subgroups are socially-meaningful for members.

The data come from a large utility with 2,685 managers. At the time the data were collected, the organization was responding to changes in its competitive environment. Company executives instituted reductions-in-force for employees they thought could not adjust to the new environment. They also hired a group of younger people with more education and placed them in higher level positions than had been traditional for entry level employees. These changes altered what had been traditional, life-long managerial careers. With this historical background, it seemed likely that these changes would influence the organizational faultlines we observed.

Demographic data on this population were obtained from the firm. A 20% systematic, stratified sample (N=537) received a survey that, among other questions, requested a list of names of the people each subject knows, the totality of which represent his or her organizational reference group (Lawrence, 2006). Survey results were received from 77% of subjects in the sampling frame (N=411). This study includes only subjects whose organizational reference groups included both close and distant associations, which reduces the sample to 358. We used this reduced sample because we are interested in faultlines defining relatively large social structures and this insures that each organizational reference group includes diverse associations. Of the 53 subjects dropped, 42 identified no close work associations, one identified no distant work associations and ten identified no known others. We compared this reduced sample to the population and found no significant differences (gender: X^2=0.25, *p*=0.62; ethnicity: X^2=0.38, *p*=0.94; age: *t*=-1.36, *p*=0.17; organizational tenure: *t*=0.30, *p*=0.77; education: *t*=0.14, *p*=0.89; career level: *t*=0.09, *p*=0.93).

Individual-Level Variables

Individuals' demographic attributes. Data on individuals' gender, ethnicity (White, Black, Hispanic and Asian), age, organizational tenure, education and career level were obtained for the population of managers (N=2,685) from company records.

Organizational reference groups. Each subject's organizational reference group comprises the set of known others he or she identified on the survey. Names were solicited by asking subjects to "copy the names of the employees you know." As is common in ego network surveys, a complete list of the 2,685 managers was provided to aid recall. Subjects provided an average of 50 names each. Although it seems likely that subjects know more than fifty people in the organization, this represents several times the number of names generated by the average ego network survey, which includes around eight (Lawrence, 2006; Marin, 2004). These lists were connected to company records, a process that involved matching around 20,000 names. As a result, although organizational reference group data are only available for sample subjects, attribute data were available for all members of all organizational reference groups.

Work-related communication contacts. After subjects completed their list of known others, they were asked with which of these others they discuss general work issues. On average, subjects engage in such discussions with 59% of their organizational reference group's members. The within- and between-subgroup measures of work-related contacts are computed after the LCCA identifies discrete subgroups. A subject's within-subgroup contacts include those subjects in his or her own subgroup with whom he or she indicates work-related contact. A subject's between-subgroup contacts include those work-related contacts who fall within other subgroups. Work-related communication contacts are asymmetric. An individual listed as a contact by Subject A is not necessarily listed as a contact by Subject B. Consequently, Subgroup A's perceived contacts with Subgroup B are not necessarily the same as Subgroup B's perceived contacts with Subgroup A.

Group-Level Variables

Subgroups are identified using the most likely class membership of subjects' organizational reference groups as assessed by LCCA. Each class identifies a subgroup and each subgroup contains subjects whose organizational reference groups are similar in composition within the subgroup and different in composition between subgroups. Composition is assessed using the six demographic attributes above. Although the identification of subgroups is based on demographic attributes for all members of subjects' organizational reference groups, only survey subjects receive subgroup assignments.

Faultlines are the individual attributes that describe the boundaries between subgroups. Similar to existing methods, the data used here begin with specific individual attributes that are assumed to be significant based on previous research. However, in contrast to extent research, the relative contribution of these attributes to faultline identification is assessed after the LCCA analysis. The contribution of each attribute is assessed qualitatively by examining which attributes seem to play the largest role in the distinctions between subgroups.

Table 2 provides descriptive statistics of all individual-level variables, including means, standard deviations and correlations.

Table 2 About Here

Results

We used LCCA (Mplus version 5.21, see L. K. Muthen & Muthen, 1998-2008) to identify subgroups within the population of subjects' organizational reference groups. We evaluated seven models, ranging from two classes to eight, and selected a five-class model as most appropriate for describing organizational faultlines. Below, we describe the LCCA procedure, provide a description of the faultline results, and test whether the attributes of the resulting subgroups provide distinctive, subgroup information. Finally, we explore the criterion-related validity of these organizational faultlines by examining within- and between-subgroup work-related communication contacts.

LCCA Procedure

The most important part of conducting LCCA is identifying the appropriate number of classes. Model comparisons and the selection of a final latent class structure are accomplished on both statistical and substantive grounds (see B. O. Muthen, 2003), with an interest towards model parsimony.

Statistical Assessment. Models with different numbers of classes are not nested and therefore incomparable using traditional fit statistics. Nyland, Asparouhov, and B. O. Muthen (2007) recommend comparing models with the Bayesian Information Criterion (BIC). The BIC is a metric of both model parsimony and fit to the data. It is derived as a function of a model's chi-square value, the number of model parameters, and sample size, where better model fit is indicated by lower BIC values (Schwartz, 1978). Although it is possible to bootstrap the likelihood ratio test when assessing fit, the BIC recovers the correct number of classes more frequently in certain cases and in all cases requires less computing time.

A second statistical method of model evaluation involves entropy values, which indicate the quality of classification. Entropy values range between 0.0 and 1.0 and are a function of the average posterior probabilities of class membership across classes. If individuals have equal posterior probabilities of membership across all classes, this means that they cannot be meaningfully assigned to a single class, which makes classification quality low. Alternately, if individuals have a high probability of membership in a single class, while having a low probability of membership in all other classes, they are easily assigned to a single class, which makes classification quality high. As noted by L. K. Muthen and B. O. Muthen (2000, p. 887), "The average posterior probability for each class for individuals whose highest probability is for that class should be considerably higher than the average posterior probabilities for the other classes for those individuals." Entropy measures the extent to which this is true, where higher values indicate better classification. B. O. Muthen (2004) suggests that values above 0.80 are acceptable.

Substantive Assessment. In addition to these quantitative methods of assessment, we substantively evaluated each model in three ways (for discussion see B. O. Muthen, 2003). Our theoretical interest was to identify structure in individuals' broad social frames of reference at work rather than just their work groups. As a result, we first examined the distribution of the number of people in each class. If adding an additional class to a model creates classes with very small numbers, then by definition that class seems unlikely to provide important information about organization-level faultlines. Second, we examined the magnitude of differences between classes along the observed variables. If adding an additional class to a model creates a class that is not substantially different from another class in terms of the observed variables, then it is also unlikely to be relevant. Finally, we evaluated the extent to which each class captures a subgroup within the organization that appears meaningfully distinct given the organization's population demography. This was done by comparing the characteristics of each class of individuals to the characteristics of the organization as a whole.

Results of LCCA Faultline Analysis

Model Comparisons. Models with between two- and eight-classes were estimated and their BIC and entropy values compared. As shown in Figure 1, BIC values decrease until an eighth class is added, at which point the level of decrease appears marginal. Given that decrements in BIC values appear to attenuate between the five- and eight-class solutions, as well as the fact that entropy values appear acceptable for all models, we moved on to substantive assessment.

Figure 1 About Here

We examined the number of subjects assigned to each class, with particular attention to the smallest. The smallest class in the five-class solution includes 31 subjects. The smallest in the six- through eight-class solutions drops to 14, which seems small for identifying the larger informal social structure that is of interest. We then compared the means of the six observed variables across models. The profiles are quite similar. While the number of classes increases, the attributes that define the most important faultlines remain. Thus, for instance, all five classes of the five-class solution are easily recognizable in the six-class solution. The only descriptive difference is that the six-class solution separates out a small group of fourteen subjects. Finally, we examined the attribute profiles for the models. The five-class solution appears to capture the organization's employment history (see descriptions below), as do the six- to eight-class solutions but with smaller class sizes. These criteria suggest that the tradeoff of slightly better fit to the data for models with six or more classes is not justified by our theoretical interest in large social structures or model parsimony. Therefore, we opted to retain the five-class solution.

Identifying Organizational Faultlines

After inductively deriving the five-class model, we describe the organizational faultlines these classes identify (see Table 3). In this organization, all individual attributes included in the analysis contribute to faultlines, with the exception of Black and Hispanic. A univariate anova across the five subgroups for every attribute shows that, with the two exceptions just mentioned, each plays a significant role in distinguishing among subgroups. Given the significance of these global *F* tests, mean comparisons of subgroup pairs were performed to examine which attributes play the most consistent roles in identifying faultlines. Subgroup pairs were first tested for equality of variances and mean comparison tests were adjusted accordingly.

Table 3 About Here

There are 10 possible subgroup pairs for each attribute: the attribute's value in subgroup 1 compared with its value in subgroup 2; the attribute's value in subgroup 1 compared with its value in subgroup 3 and so forth. Of the 10 possible comparisons, career level plays the most consistent role as a faultline, showing significant differences in all 10 comparisons. Career level is followed by age, organizational tenure and education, each showing significant differences in nine of ten. Asian and gender follow, with Asian showing seven and gender showing six significant differences. White appears to play the smallest role, showing significant differences in only four of ten. Interestingly, ethnic groups do not seem to play a substantial role in these organizational faultlines. Three of the four ethnic groups, White, Black and Hispanic, play little significant role in identifying large informal social structures.

The Five Subgroups

Given these faultline identifications, we named each subgroup based on a qualitative assessment of its main characteristics, those that distinguish it both from other subgroups and from the population. While focusing on statistical differences, subgroup names are also informed by knowledge of the firm's history and recent changes.

The first subgroup is named the "Middle-timers." The subjects in this subgroup seem somewhat average relative to other groups. Their ethnic composition, average education, and average career level are close to the population, although they are slightly younger, with lower tenure, and more likely to be women. The second subgroup is named the "Old-timers." These subjects have been around a long time, with high organizational tenure, low career levels, and other attributes that reflect the history of the firm, such as a low proportion of women, low levels of education, and non-Asian ethnic background. The third subgroup is named the "Fast Track Men." These subjects hold the highest career levels and educational credentials of any subgroup as well as the lowest proportion of women. They are young and relatively recent hires.

The fourth subgroup is "High Level Old-timers." These subjects reflect their "Old-timer" counterparts, with the exception that they include more women, have more education, and hold higher career level positions. The final subgroup is the "Asian Women Newcomers." This is the only subgroup whose main characteristic appears to be the ethnicity of its members. Seventy-four percent of the members of this subgroup are Asian. The highest proportion of Asians in any other subgroup is 20.1% for the Fast Track Men. This fifth subgroup also holds the highest proportion of women: 54.8% compared with 45.2% for the Middle-timers, which has the next highest proportion women. The Asian Women Newcomers have the lowest organizational tenure of any subgroup and appear to have been hired at relatively low career levels.

Criterion-Related Validity of Faultlines

If these faultlines are valid then the subgroups they create should be socially-meaningful to subgroup members. Our criterion-related assessment of meaningfulness is the extent to which subjects have higher numbers of within-subgroup than between-subgroup work-related communication contacts. The results in Table 4 support this assessment. Subjects consistently evidence more work-related contacts in their own subgroup than in other subgroups.

Table 4 About Here

Middle-timers appear the least isolated subgroup. Thirty-eight percent of their work-related contacts are with other Middle-timers. However, almost equal to these within-subgroup associations, 32% of their contacts are with High Level Old-timers. Their remaining contacts are distributed in roughly equal numbers across the other three subgroups. Old-timers are one of the two most isolated subgroups. Sixty-six percent of their work-related contacts are also Old-timers. Twenty-six percent of their remaining contacts are High Level Old-timers. Old Timers have few contacts who are Middle-timers or Fast Track Men and none who are Asian Women Newcomers.

Consistent with these results, the work-related communication contacts of Fast Track Men are dominated by within-subgroup associations. Fifty-percent of their contacts are in their own subgroup. The majority of their remaining contacts are split between Middle-timers and High Level Old-timers, 24% and 17% respectively. The work-related contacts of High Level Old-timers are similarly distributed among three subgroups. Forty-eight percent of their contacts are within subgroup, 23% are with Old-timers and 21% are with Middle-timers. Similar to the Old-timers, High Level Old-timers have few work-related contacts who are Fast Track Men or Asian Women Newcomers. Finally, Asian Women Newcomers are the second relatively isolated subgroup. Fifty-six percent of their work-related contacts are within-subgroup and 35% of their remaining contacts are with Middle-timers. Asian Women Newcomers have no work-related contacts with Old-timers and few who are either Fast Track Men or High Level Old-timers.

These work-related communication contact results support the LCCA subgroup identifications by showing strong career level, age, organizational tenure, education, and career level faultlines. The majority of Middle-timers' between-subgroup contacts have higher career levels and longer organizational tenure. The majority of Old-timers' between-subgroup contacts are more likely to be women and hold higher career levels and higher education. The majority of Fast Track Men's between-subgroup contacts are more likely to be women and have lower career levels, higher organizational tenure, and lower education. The majority of High Level Old-timers between-subgroup contacts are others with lower career levels. The majority of Asian Women Newcomers' between-subgroup contacts are more likely to be other women and more likely to be White. Interestingly, the two subgroups that are least often selected as work-related communication contacts are the Fast Track Men and the Asian Women Newcomers. These subgroups have the lowest average age, the lowest average organizational tenure and the highest average education.

Discussion

This study explores the use of LCCA for examining organizational faultlines, alignments of individual attributes that define relatively large, meaningful subgroups conditioned by how individuals perceive their organization. Organizational faultlines are conceptually similar to the faultlines currently being studied in small groups. The faultline concept is appealing because it recognizes a multiple attribute approach to organizational demography: individuals' demographic attributes are interdependent both empirically and through the social meaning they acquire. Multiple attribute theories exist in both psychology and sociology, some dating back to the 1970s. Examples include McGuire and Pawawer-Singer's (1976) discussions regarding the effects of attribute distinctiveness on the self-concept and Blau's (1977) consideration of how the positive correlations among attributes, or consolidation, in a social system limit the complexity of its social structure.

However, research has been limited by empirical measurement. Examining additive effects of multiple attributes is possible by summing explained variances. Examining independent explained variance is possible by examining the significance of partial correlation coefficients, assessed using incremental R^2. Examining interdependence is possible using interaction terms. However, all these techniques focus on understanding what happens to the individual rather than on how multiple attributes produce group-level phenomena. Existing faultline measures, such as *Fau* and *FSL*, are designed for examining a few alignments and, at present, become unwieldy or unusable when the number of attributes or the group is large. Moreover, they limit simultaneous analysis of attributes holding different underlying distributional properties without a transformation.

Given these difficulties of matching theory and measurement, this paper proposed and presented an empirical example of LCCA as an alternative, a complement rather than a replacement, for existing faultline methods. LCCA poses several advantages over existing measures. LCCA can be used with a relatively unlimited number of attributes, subject to model convergence issues. Further, it allows inclusion of attributes measured using nominal, categorical and continuous variables in the same analysis without additional categorization or transformation.

Finally, it induces the relative contribution of attributes to the identified subgroups. Both existing faultline measures and LCCA begin with a selected set of attributes. However, group faultline measures use theory and past empirical research to infer attribute salience. For instance, Rico et al. (2007) selected levels of conscientiousness and educational background for a laboratory study of group faultlines because previous research shows that these two attributes contribute to team performance. This is perfectly reasonable. However, LCCA provides an additional level of inquiry. It begins with similar hypotheses about which attributes are important. But it also induces which attributes from the selected set produce the most consistent contribution to faultlines. It does not assume a priori that all selected attributes are equally salient.

The study presented here identified organizational faultlines in a large company. Faultlines were identified with a LCCA of subjects' organizational reference groups. Each reference group includes a subject's known others, including many with whom he or she has little or no communication. These groups thus provide a broad view of the others in the organization of whom each subject is aware, the picture he or she has in mind when considering "what kind of people work here?" The analysis was performed using the attributes of these reference groups. Six individual demographic attributes were selected for analysis, each shown to be salient to individuals in previous research.

The LCCA resulted in a five-class solution. The attributes of the subjects classified in each of the five subgroups differ statistically along each of the six attributes. A qualitative assessment of these differences consistent with company history suggested the following descriptions: Middle-timers, Old-timers, Fast Track men, High Level Old-timers and Asian women newcomers. In general, the hierarchical attributes, including age, organizational tenure, education and career level, contributed more than the nominal attributes, including gender and ethnicity, to the organizational faultlines defining subgroups. The meaning subjects attach to these subgroups was assessed using their work-related communication contacts. As expected, the number of subjects' work-related contacts was greater within- than between-subgroups. This provides a measure of criterion-related validity for the identified faultlines.

Although these results represent a single organization, we do not expect the these faultlines to generalize. At least in this organization, "history" plays a large role in informal social structure. This suggests that organizations with different histories may, independent of similar attribute alignments, evoke different perceptions of organizational faultlines. Thus, although all relatively large organizations seem likely to have faultlines, their attributes may differ depending on the organization's size, demography and employment histories. For instance, employees in a medium-sized start-up have had less time than the employees in this organization to develop strong bonds. These bonds have less to do with individuals' friendships, close working relationships, or liking than they do with the understandings that result from many years of shared organizational experiences. For instance, "baby boomers" or those who "lived through the Great Depression" are broad groups of people who may not know one another, but who share a social identity that cannot be acquired. If this inference is generalizable, it also suggests a caution in using laboratory studies for faultline research. Status differences such as level or education can be primed, but it may be difficult to prime "years of shared experience."

Future Theoretical Directions for LCCA

This last limitation suggests several future studies. For instance, at what size do organizational faultlines cease to have meaning for subgroups? The approach here assumes that meaningful informal social structure occurs at subgroup sizes bigger than work groups and smaller than large organizations. It seems likely that there is some balance between the size of the group and the salience of the attributes involved that produces a “maximum” subgroup size. Once subgroups get too large, they probably represent one of two extremes: either they become representative of the population or they become so different that faultline analysis is required on only a few attributes. Li and Hambrick (2005) present a good example of this. All of the top management teams they studied were characterized by two factions, the locals and the expatriates. Thus it was unnecessary to include this attribute in the faultline analysis.

A method of course is only as good as the theories for which it is useful. LCCA seems particularly appropriate for theories in which the central concept involves a profile, such as a profile of individuals’ personality traits, a profile of high versus low-performing groups or a profile of the network attributes of industries. Thus, it is appropriate for theories at several levels of analysis, including those that involve groups of groups rather than groups of individuals.

Zyphur (forthcoming) provides an example of the profile approach. Psychologists have long studied the relationship between personality and job performance studies. The assumption behind these studies is that there is something about an individual’s personality that results in higher or lower performance. However, using LCCA one might pose the question as to whether there are interdependencies among performance and personality traits that produce distinct subgroups that illuminate various profiles of personality and performance. For example, one class might include individuals with high job performance, low scores on Openness to Experience and high scores on Agreeableness. Another class might also include individuals with high job performance, but involve low scores on Openness to Experience and Agreeableness. These would not be distinguishable as different profiles using regression analysis without an unwieldy number of interaction terms.

At a different level of analysis, the results presented here could be turned inside out to focus not on the distinctions among the individuals but rather on the distinctions among the subgroups. In most faultline studies, subgroups are derived from demographic alignments among individuals. In this study, subgroups result from demographic alignments among organizational reference groups. Thus, these subgroups represent groups of groups: subgroups of organizational reference groups that are similar in composition within a class and different across classes. It seems possible and perhaps even likely that the attributes of the individuals who fall into these subgroups differ from the attributes of the subgroups themselves. An individual does not need to be young to belong to a subgroup or neighborhood characterized by young people. Focusing on the attributes of these groups of groups might provide a new perspective on what emergent social structures look like in an organization.

These examples at different levels of analysis suggest many possibilities for using LCCA. The method facilitates study of theoretical questions that have been difficult to explore because of empirical issues, such as the organizational faultlines presented here. Moreover, it encourages theory-building by providing researchers with an alternate strategy for conceptualizing the interrelations among variables that may challenge existing thought.

References

Ashforth, B. E., & Johnson, S. A. (2001). Which hat to wear? The relative salience of multiple identities in organizational contexts. In M. A. Hogg & D. J. Terry (Eds.), *Social Identity Processes in Organizational Contexts* (pp. 31-48). Philadelphia: Psychology Press.

Barkema, H. G., & Shvyrkov, O. (2007). Does top management team diversity promote or hamper foreign expansion? *Strategic Management Journal, 28*, 663-680.

Bezrukova, K., Jehn, K. A., Zanutto, E. L., & Thatcher, S. M. B. (2009). Do Workgroup Faultlines Help or Hurt? A Moderated Model of Faultlines, Team Identification, and Group Performance. *Organization Science, 20*(1), 35-50. doi: 10.1287/orsc.1080.0379

Blau, P. M. (1977). *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: The Free Press.

Browne, I., & Misra, J. (2003). The Intersection of gender and race in the labor market. *Annual Review of Sociology, 29*, 487-513.

Cannella, A. A., Park, J.-H., & Lee, H.-U. (2008). Top management team functional background diversity and firm performance: Examining the roles of team member colocation and environmental uncertainty. [Article]. *Academy of Management Journal, 51*(4), 768-784.

Castilla, E. J. (2008). Gender, Race, and Meritocracy in Organizational Careers. *American Journal of Sociology, 113*(6), 1479-1526.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral sciences (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associated.

Cohen, L. E., Broschak, J. P., & Haveman, H. A. (1998). And then there were more? The effect of organizational sex composition on the hiring and promotion of managers. *American Sociological Review, 63*(5), 711-727.

DiStefano, C., & Kamphaus, R. W. (2006). Investigating subtypes of child development: A comparison of cluster analysis and latent cluster analysis in typology creation. *Educational and Psychological Measurement, 66*(5), 788-794.

Elvira, M. M., & Cohen, L. E. (2001). Location Matters: A Cross-Level Analysis of the Effects of Organizational Sex Composition on Turnover. *The Academy of Management Journal, 44*(3), 591-605.

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling, 8*, 128-141.

Flache, A., & Mas, M. (2008). How to get the timing right. A computational model of the effects of the timing of contacts on team cohesion in demographically diverse teams.

Computational and Mathematical Organization Theory, 14(1), 23-51. doi: 10.1007/s10588-008-9019-1

Hamaker, E. L., Dolan, C. F., & Molenaar, P. C. M. (2003). ARMA-based SEM when the number of time points T exceeds the number of cases N: Raw data maximum likelihood. *Structural Equation Modeling, 10*, 352-379.

Homan, A. C., Hollenbeck, J. R., Humphrey, S. E., Van Knippenberg, D., Ilgen, D. R., & Van Kleef, G. A. (2008). Facing differences with an open mind: Openness to experience, salience of intragroup differences, and performance of diverse work groups. *Academy of Management Journal, 51*(6), 1204-1222.

Homan, A. C., van Knippenberg, D., Van Kleef, G. A., & De Dreu, C. K. W. (2007). Bridging faultlines by valuing diversity: Diversity beliefs, information elaboration, and performance in diverse work groups. *Journal of Applied Psychology, 92*(5), 1189-1199.

Klauer, K. C., Ehrenberg, K., & Wegener, I. (2003). Crossed categorization and stereotyping: Structural analyses, effect patterns, and dissociative effects of context relevance. *Journal of Experimental Social Psychology, 39*, 332-354.

Lau, D. C., & Murnighan, J. K. (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review, 23*(2), 325-340.

Lau, D. C., & Murnighan, J. K. (2005). Interactions within groups and subgroups: The effects of demographic faultlines. *Academy of Management Journal, 48*(4), 645-659.

Lawrence, B. S. (2006). Organizational reference groups: A missing perspective on social context. *Organization Science, 17*, 80-100.

Li, J., & Hambrick, D. C. (2005). Factional groups: A new vantage on demographic faultlines, conflict, and disintegration in work teams. *Academy of Management Journal, 48*(5), 794-813.

Marin, A. (2004). Are respondents more likely to list alters with certain characteristics? Implications for name generator data. *Social Networks, 26*, 289-307.

McGuire, W. J., McGuire, C. V., & Winton, W. (1979). Effects of household sex composition on the salience of one's gender in the spontaneous self-concept. *Journal of Experimental Social Psychology, 15*, 77-90.

McGuire, W. J., & Pawawer-Singer, A. (1976). Trait salience in the spontaneous self-concept. *Journal of Personality and Social Psychology, 33*, 743-754.

Mehra, A., Kilduff, M., & Brass, D. J. (1998). At the margins: A distinctiveness approach to the social identity and social networks of underrepresented groups. *Academy of Management Journal, 41*, 441-452.

Molleman, E. (2005). Diversity in demographic characteristics, abilities and personality traits: Do faultlines affect team functioning? *Group Decision and Negotiation, 14*, 173-193.

Mortensen, M. (2008). *Fuzzy teams: Why do teams disagree on their membership, and what does it mean?* Working paper. MIT. Boston, MA.

Muthen, B. O. (2003). Statistical and substantive checking in growth mixture modeling. *Psychological Methods, 8*, 369-377.

Muthen, B. O. (2004). Latent variable analysis: Growth mixture modeling In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthen, L. K., & Muthen, B. O. (1998-2008). *Mplus Users Guide* (3rd Edition ed.). Los Angeles: Authors.

Muthen, L. K., & Muthen, B. O. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882-891.

Nyland, K., Asparouhov, T., & Muthen, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation. *Structural Equation Modeling, 14*, 535-569.

Pearsall, M. J., Ellis, A. P. J., & Evans, J. M. (2008). Unlocking the effects of gender faultlines on team creativity: Is activation the key? *Journal of Applied Psychology, 93*(1), 225-234. doi: 10.1037/0021-9010.93.1.225

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the Black Box: An Analysis of Work Group Diversity, Conflict, and Performance. *Administrative Science Quarterly, 44*(1), 1-28.

Polzer, J. T., Crisp, C. B., Jarvenpaa, S. L., & Kim, J. W. (2006). Extending the faultline model to geographically dispersed teams: How colocated subgroups can impair group functioning. *Academy of Management Journal, 49*(4), 679-692.

Rico, R., Molleman, E., Sanchez-Manzanares, M., & Van der Vegt, G. S. (2007). The effects of diversity faultines and team task autonomy on decision quality and social integration. *Journal of Management, 33*(1), 111-132.

Sawyer, J. E., Houlette, M. A., & Yeagley, E. L. (2006). Decision performance and diversity structure: Comparing faultlines in convergent, crosscut, and racially homogeneous groups. *Organizational Behavior and Human Decision Processes, 99*, 1-15.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464.

Shaw, J. B. (2004). The development and analysis of a measure of group faultlines. *Organizational Research Methods, 7*(1), 66-100.

Thatcher, S. M. B., Jehn, K. A., & Zanutto, E. (2003). Cracks in diversity research: The effects of diversity faultlines on conflict and performance. *Group Decision and Negotiation, 12*, 217-241.

Urada, D., Stenstrom, D. M., & Miller, N. (2007). Crossed categorization beyond the two-group model. *Journal of Personality & Social Psychology, 92*(4), 649-664.

Vescio, T. K., Hewstone, M., Crisp, R. J., & Rubin, J. M. (1999). Perceiving and responding to multiply categorizable individuals: Cognitive and affective intergroup bias. In D. Abrams & M. A. Hogg (Eds.), *Social Identity and Social Cognition* (pp. 111-140). Malden, MA: Blackwell.

Zyphur, M. (forthcoming). When mindsets collide: Switching analytical mindsets ot advance organizational science. *Academy of Management Review*.

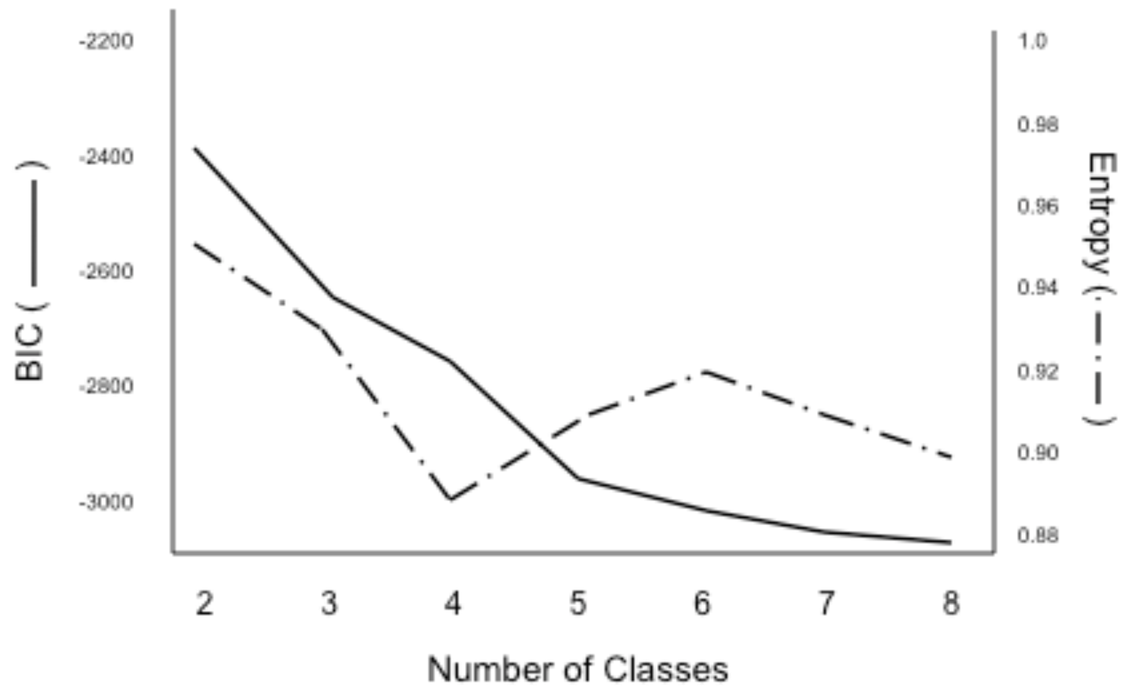

***Figure 1*.** **BIC and entropy values for two- through eight-class models. Lower values indicate better fit.**

TABLE 1
Methods of Faultline Identification

Empirical Method	Size of Social System	Number of Subgroups Possible	Number of Attributes	Types of Variables			Requires Transforming/ Weighting Variables with Different Distributions
				Nominal	Categorical	Continuous	
LCCA	Any size	Limited by degrees of freedom	Limited by convergence issues	Yes	Yes	Yes	No
Fau	Small only[a]	2	Exponentially more complex as number of attributes increase	No	Yes	Yes	Yes
FLS	Small only	No actual subgroup identification	Exponentially more complex as number of attributes increase	Yes	Yes	No	Yes
d	Any size	No actual subgroup identification	Unlimited, easy to collapse across multiple variables	No	Yes	Yes	Yes

LCCA = latent class cluster analysis; *FLS* = faultline score.
[a] Recent revision designed to increase size (K. Bezrukova, personal communication, August 11, 2009).

TABLE 2
Means, Standard Deviations and Correlation Matrix (N=358)

Individual Attributes	X	SD	1	2	3	4	5	6	7	8	9	10
1. Gender	0.30	0.46										
2. White	0.62	0.49	-0.16									
3. Black	0.09	0.29	0.05	-0.40								
4. Hispanic	0.16	0.37	0.01	-0.56	-0.14							
5. Asian	0.13	0.33	0.19	-0.49	-0.12	-0.17						
6. Age	42.99	8.32	-0.12	0.16	0.07	-0.11	-0.18					
7. Organizational Tenure	17.20	9.72	-0.19	0.19	0.06	-0.03	-0.30	0.81				
8. Education	5.75	1.07	0.03	-0.07	0.19	-0.11	0.28	-0.27	-0.39			
9. Career Level	4.55	7.55	-0.21	0.13	-0.07	-0.15	-0.04	0.15	0.09	0.34		
Organizational Reference Group Attributes												
10. Proportion Women	0.31	0.18	0.68	-0.20	0.08	0.01	0.22	-0.15	-0.28	0.19	-0.12	
11. Proportion White	0.63	0.16	-0.33	0.56	-0.28	-0.21	-0.35	0.27	0.36	-0.14	0.18	-0.51
12. Proportion Black	0.10	0.08	0.19	-0.36	0.62	0.07	-0.07	0.08	0.06	-0.14	-0.17	0.27
13. Proportion Hispanic	0.15	0.08	0.07	-0.23	0.04	0.47	-0.22	0.01	0.13	-0.25	-0.21	0.02
14. Proportion Asian	0.12	0.13	0.25	-0.33	-0.06	-0.06	0.61	-0.39	-0.55	0.40	0.00	0.45
15. Average Age	44.11	3.40	-0.12	0.23	0.05	-0.01	-0.38	0.68	0.76	-0.45	0.00	-0.30
16. Average Org Tenure	18.64	4.76	-0.18	0.25	0.05	0.02	-0.43	0.62	0.77	-0.50	-0.06	-0.37
17. Average Education	2.68	0.44	0.18	-0.17	-0.03	-0.05	0.32	-0.28	-0.46	0.56	0.40	0.42
18. Average Career Level	11.75	1.06	-0.11	0.15	-0.11	-0.13	-0.00	0.15	0.05	0.37	0.68	-0.06
19. Work-Related Communication Contacts	29.82	16.17	-0.05	0.24	-0.09	-0.13	-0.13	0.03	0.01	0.18	0.28	-0.02

$r > 0.104 = p < 0.05$. Women, White, Black, Hispanic, & Asian = 1.

TABLE 2 (continued)
Means, Standard Deviations and Correlation Matrix (N=358)

Individual Attributes	11	12	13	14	15	16	17	18
1. Gender								
2. White								
3. Black								
4. Hispanic								
5. Asian								
6. Age								
7. Organizational Tenure								
8. Education								
9. Career Level								
Organizational Reference Group Attributes								
10. Proportion Women								
11. Proportion White								
12. Proportion Black	-0.54							
13. Proportion Hispanic	-0.41	0.20						
14. Proportion Asian	-0.66	-0.09	-0.21					
15. Average Age	0.44	0.10	0.16	-0.69				
16. Average Org Tenure	0.45	0.09	0.24	-0.76	0.95			
17. Average Education	-0.27	-0.13	-0.32	0.60	-0.59	-0.70		
18. Average Career Level	0.30	-0.26	-0.36	-0.00	-0.00	-0.10	0.59	
19. Work-Related Communication Contacts	0.21	-0.17	-0.13	-0.08	0.02	0.01	0.12	0.27

$r > 0.104 = p < 0.05$. Women, White, Black, Hispanic, & Asian = 1.

TABLE 3
Description of Organizational Subgroups Defined by LCCA Faultlines (N=358)

Organizational Subgroups	N	Organizational Subgroup Attributes: Means and Proportions								
		Gender	White	Black	Hispanic	Asian	Age	Org Tenure	Education	Career Level
(1) Middle-timers	73	0.452	0.575	0.137	0.164	0.123	38.75	11.95	2.59	7.69
(2) Old-timers	136	0.162	0.691	0.103	0.191	0.015	45.19	21.50	1.92	6.33
(3) Fast Track Men	34	0.147	0.647	0.029	0.118	0.206	35.12	6.24	3.38	9.53
(4) High Level Old-timers	84	0.381	0.714	0.071	0.155	0.048	49.07	23.71	2.45	8.95
(5) Asian Women Newcomers	31	0.548	0.161	0.032	0.065	0.742	35.45	5.10	3.00	6.61
F		9.74^{***}	9.44^{***}	1.36^{ns}	0.89^{ns}	49.72^{***}	50.00^{***}	86.94^{***}	41.42^{***}	19.54^{***}
Subgroup comparisons		12, 13, 24, 25, 34, 35	15, 25, 35, 45	13, 15	25	12, 15, 23, 25, 34, 35, 45	12, 13, 14, 15, 23, 24, 25, 34, 45	12, 13, 14, 15, 23, 24, 25, 34, 45	12, 13, 15, 23, 24, 25, 34, 35, 36	12, 13, 14, 15, 23, 24, 25, 34, 35, 45
Population Means	2685	0.316	0.617	0.098	0.159	0.121	43.59	17.05	2.71	7.55

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
Subgroup comparisons: ^{xy} = Mean of Subgroup X differs from that of Subgroup Y, $p < 0.05$.
LCCA = latent class cluster analysis.

TABLE 4
Work-Related Communication Contacts Within- and Between-Organizational Subgroups (N=358)

		Average Percent of Work-Related Communication Contacts With:						
Organizational Subgroups	N	(1) Middle-timers	(2) Old-timers	(3) Fast Track Men	(4) High Level Old-timers	(5) Asian Women Newcomers	Total N	F
(1) Middle-timers	73	**0.38**	0.11	0.09	0.32	0.10	313	39.49 ***
(2) Old-timers	136	0.07	**0.66**	0.01	0.26	0.00[a]	535	1363.30***
(3) Fast Track Men	34	0.24	0.03	**0.50**	0.17	0.06	115	104.67***
(4) High Level Old-timers	84	0.21	0.23	0.05	**0.48**	0.02[a]	399	148.05***
(5) Asian Women Newcomers	31	0.35	0.00[a]	0.04	0.05	**0.56**	103	43.85***

*** $p < 0.001$.

Within-subgroup communications in bold.

[a] Subgroup dropped from global multivariate test because matrix not positive definite.