Identifying Organizational Faultlines with Latent Class Cluster Analysis

Barbara S. Lawrence
Michael J. Zyphur

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

January 16, 2010

ABSTRACT

Faultline theory proposes that when the distribution of individuals' attributes in groups are aligned 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 appealing, empirical methods have been difficult. The most commonly used, *Fau* and *FLS*, have several limitations, for instance difficulty with integrating nominal, categorical, and continuous variables. This paper proposes latent class cluster analysis (LCCA) as an additional analytical tool. After reviewing the literature involving interdependent attributes, 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, 1999p. 230), turnover (Elvira & Cohen, 2001), performance (Cannella, Park, & Lee, 2008), corporate foreign investment (Barkema & Shvyrkov, 2007), career mobility (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.

Faultlines (Lau & Murnighan, 1998) "are hypothetical dividing lines" (p. 328) within a group defined by the alignment of members' demographic attributes. These dividing lines produce subgroups that place structural constraints on intragroup relationships by accentuating members' within-subgroup similarities and between-subgroup differences. Moreover, they index salient contextual information because individuals often associate demographic attributes with important work outcomes such as recognition, salary or promotion (Kanter, 1977b; Ridgeway, 1991). When attribute distributions coincide, salient meanings coincide. This reinforces group members' awareness of the aligned attributes, making faultlines cognitively-accessible[1] and consequential boundaries. 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, called faultline strength by others (Zanutto, Bezrukova, & Jehn, in press), as measured by

---

[1] Cognitively-accessible is "the readiness with which stored knowledge can be used" (Fiske & Taylor, 1991, p. 247).

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. Faultlines are defined by the joint distribution of several attributes rather than the individual distributions of each. This interdependence distinguishes faultline theory from most work on organizational demography, which often assumes attribute independence and measures additive or averaged distributional effects (e.g., Polzer, Milton, & Swann, 2002; Urada, Stenstrom, & Miller, 2007). Such independent effects are easy to study with standard techniques such as regression. However, attribute interdependence presents measurement issues that have limited the development of faultline research and theory.

The current study uses latent class cluster analysis (LCCA) to extend the empirical and theoretical territory of faultline research. LCCA has several useful properties. First, consistent with current faultline studies, it can be applied to small groups when each group is considered a population. It can also be used in large groups even when the group is considered a sample. Second, it is scale-invariant and insensitive to linear transformations of observed variables. It accommodates non-continuous variables with a wide variety of underlying distributions without requiring transformation by relying on the appropriate linking functions and joint distributions. Current methods used to identify faultlines are similar to cluster-analysis and, as is common with these methods, authors recommend transforming variables to account for sensitivity to the scaling of observed variables (e.g., Zanutto, et al., in press).

Third, LCCA is model-based and probabilistic. A model is specified and parameters are estimated to provide the best fit to the observed data. This permits assessing the number of faultlines by comparing the fit obtained with different numbers of subgroups. LCCA also returns

model-estimated subgroup properties, such as subgroup means for observed variables. This facilitates computation of posterior probabilities of membership for each individual for each subgroup. These probabilities provide a measure of faultline strength using an entropy statistic based on model-estimated parameters and each individual's observed data. These characteristics differ from previous methods, which specify a single number of subgroups, generally two, prior to analysis. Further, previous assessments of subgroup membership and faultline strength do not indicate the uncertainty in subgroup assignment.

We capitalize on these LCCA properties to extend faultline theory by examining organizational rather than small workgroup faultlines. Organizational faultlines are attribute alignments in a large group whose boundaries are defined by the population of others that individuals know, such as those with whom they work or talk, about whom they hear or whom they merely observe. Although it is possible to study faultlines using an organization's actual demographic distributions, our interest is in the organization that emerges from members' perceptions. This is the social region between the individual and organization within which norms, values and expectations evolve. It seems probable that organizational faultlines define subgroups that become repositories of the salient information individuals collect over time to make sense out of their work experiences.

There are two important differences between workgroup and organizational faultlines. One concerns knowledge of others and their demographic attributes. Individuals in small workgroups know one another and are likely aware of one another's demographic attributes. When a group gets sufficiently large, individuals cannot know everyone. This means their awareness of demographic attributes depends on the sample of others they do know, and it seems likely these samples are non-random. Each individual has his or her own sample: an

organizational reference group (Lawrence, 2006)[2]. This suggests that a second difference between workgroup and organizational faultlines is who defines the group's boundaries. Small workgroup membership is usually formally-prescribed by authorities, perhaps by a manager or researcher. However, in large groups or organizations, the group's boundaries may not be prescribed. Instead, they may be defined by members' perceptions, which may differ from the actual population within formal organizational boundaries.

The discussion below is organized as follows. We begin with a review of existing theories that involve attribute interdependence, focusing on their similarities and differences in identifying attribute-based subgroups. This is followed by a longer evaluation of existing faultline methods, including their identification, strength and distance. We conclude with an empirical study in a large organization. We use LCCA to identify organizational faultlines and then assess their criterion-related validity using work-related communication contacts to assess within- and between-subgroup interaction.

Explaining the Effects of 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 for why attribute alignments influence behavior.

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. For instance, demographic attributes

---

[2] An organizational reference group is "the set of people an individual perceives as belonging to his or her work environment that defines the social world of work in which he or she engages, including people with whom the individual does and does not communicate and those with whom awareness is the only connection" (Lawrence, 2006, p. 80).

such as age and gender frequently become distinctive when their distribution is skewed (Kanter,

1977b). When only a few individuals are young or only a few are women, these attributes garner

attention. This increases the probability that individuals use these attributes as a basis for social

identity (Mehra, Kilduff, & Brass, 1998). 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.

Crossed-attribute categorization presents another explanation for why attribute

alignments influence behavior. 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 individuals'

perceptions of similarity increase with each added homophilous attribute. The feature detection

strategy is particularly relevant to organizational decisions that relate to performance, such as

promotions and salary. Crossed-attribute categories acquire significant meaning when they are

related to performance. This may occur through a social comparison process called related-

attributes theory (Goethals & Darley, 1977; Wheeler, 1991), which proposes that individuals

compare themselves to others whose attributes are related to the salient outcome.

Intersectionality and consolidation theory emphasize structural explanations for why

attribute alignments influence behavior. These explanations suggest that the joint distribution of

attributes 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, p. 276), which he defines as the strength of the positive correlations among attributes in a given population. As these correlations increase, the size of groups with similar attributes increases, the number of groups with similar attributes decreases and the differences between groups of similar others increase. As a result, individuals likely engage in greater within- than between-group interaction, producing both fewer cordial and fewer conflictual intragroup interactions.

Blau (1977) suggests that attributes with nominal and continuous distributions play different roles in this process. Those with nominal distributions, such as gender and ethnicity, define a social system's heterogeneity. As their alignment increases, the probability that two random individuals will belong to a group of others with similar attributes increases. Attributes with continuous distributions, such as age and organizational tenure, define a social system's inequality. As their alignment increases, the diversity of individuals' possible status differences decreases. However, in both cases, increasing alignment reduces the number of attribute-based

groups. Thus, an organization's social structure, as depicted by the correlations among its members' individual attributes, represents an important structural constraint on behavior.

Faultline theory suggests a mix of these 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. Faultline strength is the extent to which these attributes are aligned, and faultline distance is the extent to which the subgroups diverge as a result of differences across them (Zanutto et al., in press). 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 explanation involves elements from other theories of attribute interdependence. Similar to consolidation theory, faultline theory suggests that individuals' opportunities for interaction increase with the strength of attribute alignments. Similar to distinctiveness and crossed-attribute theories, it suggests that attribute alignments encourage self-categorization and social identity and that these processes facilitate the emergence of behaviorally-meaningful subgroups.

All of these approaches agree that attribute alignments tend to heighten individuals' similarities and differences, and that these similarities and differences influence interaction. The dominant explanation for this influence involves attribute salience: the higher the salience of the attributes involved, the greater their impact on interaction. However, scholars propose different reasons for why attributes are or become salient. Some suggest that attributes are salient when

they represent fundamental human characteristics, such as gender and age (Linton, 1936). Others indicate that attributes become salient when they meet individuals' needs to be both similar to and different from others (Brewer, 1991; Hogg & Terry, 2000). Still others believe attributes become salient when they provide evaluative information, such as individuals' status and authority within a given social context (Blau, 1977; Kanter, 1977a) or their acquisition of significant resources (Ridgeway, 1991). In the end, salience likely results from some combination of psychological and structural mechanisms.

Given the importance of attribute salience, current theories pay insufficient attention to the role that perceptions play in faultline effects. Some evidence exists that people are aware of faultline splits (Zanutto, et al., in press), but we know little about the extent to which such awareness shapes faultline effects. The role of perceptions also raises questions about the function of known versus ambiguous distributions. Faultlines have been studied almost exclusively in small workgroups, in which the number of group members and distribution of attributes are relatively unambiguous. Yet, as suggested by more structural theories (Blau, 1977), faultlines play a role in larger social systems. When a group gets sufficiently large, the number of group members and the distribution of attributes become somewhat ambiguous to members (see Mortensen, 2008 for an example), which likely produces greater variation in faultline perceptions. This may alter attribute salience because the group's boundaries are influenced by factors such as co-workers' attributes, geographic proximity and perceptual biases.

Organizational vs. Workgroup Faultlines

In this study, we examine whether LCCA identifies one or more sets of aligned individual attributes that define distinct and meaningful subgroups in a large organization. For the most part, existing faultline theories explain faultline effects with psychological, social

psychological or structural mechanisms as if these mechanisms are independent. However, in a large group, faultline effects may result from an intermediate social structure that emerges from their interdependence. Individuals develop shared perceptions of demographic distributions that acquire meaning through social norms and expectations (Lawrence & Tolbert, 2007). These norms and expectations both surface from and influence how individuals make sense of their organizational experiences. As a result, individuals' perceived distributions are salient. Faultlines that emerge from these perceptions likely define behaviorally-significant, informal social structure. This type of social structure may help describe and perhaps explain the evolution of organizational subcultures, whose basic ingredient is differential interaction within- and between-groups (Trice & Morand, 1991).

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- and between-subgroup interaction. However, there are several important differences.

The first involves group membership and boundaries. As noted earlier, workgroup faultline studies involve small groups of prescribed members. External authorities select members and assign them specific tasks with one another. Everyone knows everyone else and group boundaries are unambiguous. In contrast, organizational faultlines involve a large group of perceived members. Members are drawn from individuals' organizational reference groups, the broad set of others of whom each individual is aware (Lawrence, 2006). Everyone cannot know everyone else, they may or may not be required to work together and they may or may not be friends. In a large organization each individual knows a different set of others. As a result, the group's boundary differs from that of the organization: it becomes ambiguous.

These differences in membership boundaries influence the immediacy of subgroup effects. In small workgroups, behavioral consequences are immediate. Subgroup membership, the meaning it acquires for members and its implications for within- and between-subgroup interaction, are integrated into everyday work. The connections between subgroup membership, members' attributes and acquired subgroup meanings are easily observed, chronically accessible and therefore frequently reinforced. In a large group, subgroup effects are not immediate. The meanings subgroups acquire accumulate over time and likely involve salient impressions, activities or events that do not reflect day-to-day activity. Examples might include individuals' reputation in the organization, their organization-specific knowledge or their career success.

Similar to workgroup faultline effects, we expect higher within- than between-subgroup interaction. However, unlike workgroup faultlines where subgroup membership is tightly connected to members' attributes, we expect subgroup membership that is, to some extent, independent of member's attributes. Individuals may differ from the others of whom they are aware in a large organization. A young, Hispanic woman might belong to the same subgroup as an old, White man. Individuals' attributes are decoupled from subgroup attributes. Subgroup membership means only that members know others who share common attributes.

Extant Methods for Identifying and Measuring Faultlines

The current faultline literature includes several approaches to examining faultlines (Zanutto, et al., in press): those that identify faultlines in groups and, therefore, categorize individuals into subgroups; those that measure faultline strength, or the degree of attribute alignment across the members of a group; and those that measure faultline distance, or the Euclidean distance between subgroups along observed variables.

Faultline Distance

Lau & Murnighan's (1998) discussion of faultline analysis focuses on what we define here as faultline distance. 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). Lau and Murnighan's (1998) separation of categorical and continuous measures likely represents empirical limitations of existing dispersion indices rather than theoretical exclusion.

To measure faultline distance between subgroups, Bezrukova, Jehn, and Zanutto (2009) suggest taking the Euclidean distance between the centroids of the two subgroups' multivariate distributions for the attributes in question. This value, referred to as D, examines the distance between the vector of means for the variables (Molleman, 2005). It describes the actual distance between subgroups along all attributes irrespective of the amount of within- and between-subgroup variance.

Faultline Strength

The majority of faultline measures are measures of faultline strength, the extent to which members of each subgroup share similar attributes that differ from those in other subgroups. Li and Hambrick (2005) suggest this occurs when "two factions differ in their averages [along a given attribute] *and* each faction is tightly clustered around its own average" (p. 804). They measure strength by assessing the difference between subgroups along multiple variables and then dividing these differences by their total variation, similar to an η^2. They recommend standardizing the result for each variable because of scaling differences.

Thatcher, Jehn, and Zanutto (2003) measure faultline strength with *Fau*, a direct measure of the ratio of between-subgroup to total variance in a variable, collapsed across all variables. *Fau* captures the homogeneity of subgroups, that is, the cleanness of their split. In order to place variables on a common metric, Thatcher et al. (2003) recommend standardizing continuous variables and transforming discontinuous variables to make them comparable. Similar to *Fau*, Barkema and Shvyrkov (2007) measure faultline strength with the ratio of total variation to within-subgroup variation, where larger values indicate greater strength.

Another method for identifying faultline strength is outlined by Shaw (2004). Shaw bases his measure on the idea that people perceive continuous attributes in meaningfully discrete categories. He makes continuous variables discontinuous, placing group members into subgroups along each variable. Shaw's (2004) overall faultline strength score for a group (*FLS*) is then assessed by computing the internal alignment (IA) along each variable, defined as "the extent to which members within a particular subgroup are similar to one another on all other relevant variables" (p. 72). He then weights IA by an index of "cross-subgroup alignment," defined as the degree to which individuals in other categories have similar cross-category memberships. The result of this process yields faultline strength (*FLS)*.

Finally, Gibson and Vermeulen (2003) measure faultline strength by computing the degree of similarity for every possible pairing of individuals within a group. When similarity is high for two group members, they assign a large weight to the pair, and when low they assign a low weight. They then compute the standard deviation of similarity across all possible pairings. When similarity is high for some pairings, but low for others, this produces high variance in similarity scores across the pairings. This variance indicates faultlines are present, because some group member pairings show similarity, while other pairings do not.

Identifying Faultlines

A central characteristic of faultlines is that they are latent or informal boundaries, which by definition define latent or informal subgroups. These subgroups may or may not be observed by group members and researchers. In experimental contexts researchers define faultlines *a priori*, so boundary identification is not an issue (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 in applied contexts identifying faultlines is more difficult.

Existing methods for faultline identification often rely on faultline strength. In the first work to develop a method for identifying faultlines, Thatcher et al., (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) where groups may only split into two subgroups. 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, small groups frequently include only a few individuals, making more than two meaningful subgroups unlikely, and second, the computational complexity of *Fau* increases with more than one faultline.

As noted above, *Fau* is the proportion of between-subgroup variance to total variance, similar to an intra-class correlation coefficient or η^2. *Fau* is then calculated for *S* many possible group splits *g*, where

$$S = 2^{n-1} - 1 \tag{1}$$

This is similar to combinatorial cluster analysis, where all possible combinations of individuals into subgroups are examined (Aldenderfer & Blashfield, 1984). Following the computation of *Fau* for all the *g* splits, Thatcher et al. (2003) choose the group split that produces the largest overall proportion of between-subgroup to total variance, a criterion used in other cluster-analytic methods (Aldenderfer & Blashfield, 1984). This split, in which Fau_g is closest to 1.0, indicates the faultline and defines the two subgroups. *Fau* thus reflects faultline theory by suggesting that a faultline is identified by the attribute alignment that produces the greatest within-group similarities and between-group differences. The attribute alignment with the highest strength, the one that produces the most homogeneous subgroups, is the faultline.

Thatcher et al. (2003) propose a *Fau* scaling scheme that allows the simultaneous use of continuous and non-continuous attributes. This scheme assigns comparable weights to differences across people along variables with different underlying distributions and scales. When mixing nominal, categorical and continuous variables, the researcher categorizes or weights the variables in terms of their relative importance in faultline formation. This requires a number of assumptions (e.g., Bezrukova, Jehn, Zanutto, & Thatcher, 2009; Shaw, 2004).

Although *Fau* has been limited to small groups and one faultline, recent modifications facilitate its use in larger groups (K. Bezrukova, personal communication, September 24, 2009). Additionally, there is no inherent limitation to two subgroups, as the method may be used for multiple subgroups. However, just as with other combinatorial cluster analytic techniques (Aldenderfer & Blashfield, 1984), the criterion of deciding on the correct number of subgroups by examining proportions of between subgroup to total variation becomes difficult. As the number of subgroups increases, the proportion of between subgroup variance also increases. For example, if a group is partitioned into subgroups with two individuals in each subgroup, then

between-subgroup variation will be substantial in almost every case. Therefore, although *Fau*

could be used to identify subgroups when a number of subgroups is specified *a priori,* such as

the two subgroups currently used, it probably should not be used as a method for both identifying

faultlines and identifying the number of subgroups that should be modeled.

Alternatively, and similar to our suggestions here, Barkema and Shvyrkov (2007) use a

form of latent class cluster analysis to uncover the existence of faultlines in small groups.

However, their work does not describe how they used LCCA or specified their models and they

do not compare LCCA to existing faultline research and methods. In order to understand how

LCCA may be useful to faultline researchers, such elaborations are important.

Similarities and Differences Across Faultline Measures

Most faultline measures begin with the assumption that true faultlines are latent and

unknown to the researcher. For instance, although the current version of *Fau* assumes one

faultline, it also assumes that subgroup attributes are unknown prior to analysis. Shaw (2004)

suggests that faultline boundaries are always latent. Rather than identify subgroups, *FLS*

identifies a group's faultline strength based on the cumulative internal alignment of members'

attributes.

All measures of faultline strength, in one form or another, examine the extent to which

individuals align along multiple attributes in a way that separates them from other individuals

who align with one another. For example, *Fau* measures proportions of variance, which is

similar to Li and Hambrick's (2005) measure of faultline size. Another similarity is that all of

these methods require scaling or transforming observed variables in one fashion or another. For

FLS, continuous variables must be made discontinuous. Measures relying on variances generally

require scaling variables to equate them along a common scale.

Across all methods, only two identify faultlines: *Fau* and LCCA. The most commonly used method, *Fau*, currently identifies one faultline using the amount of between-group variance it creates. The fit of any model to the original data is unknown outside of what is essentially an η^2 statistic, which is similar to a fit assessment but currently 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.

LCCA and Organizational Faultlines

What Lau and Murnighan (1998) explore as a subgroup created by a faultline is conceptually identical to what has been explored elsewhere as a "latent class" derived from a LCCA (LCCA; DiStefano & Kamphaus, 2006). A latent class is a group of individuals whose attributes exhibit more homogeneity as a cluster than the known group from which they are drawn. This homogeneity among subgroup members is not directly observed. 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).

One way to conceptualize LCCA is that, just as in latent profile analysis, it is a technique wherein individuals are assumed to belong to one of *K* latent classes. Members of a sample come from different populations but are mixed in the sample. For example, a sample with older black males and younger white women might be mixed in a dataset but come from two distinct populations. Using maximum likelihood estimation, LCCA treats the properties of these classes, such as the means along observed variables for each class, as unknown and maximizes their

likelihood in relation to observed data. Additionally, the number of classes and their size are

unknown and estimated. These estimates are often the focus of researcher using LCCA to

describe relationships among members of a population. With an interest towards model

parsimony, models with different numbers of classes are compared along both statistical and

substantive grounds (see Muthen, 2003) (see B. O. Muthen, 2003) to choose a final latent class

structure. A variety of methods exist for determining class enumeration and just as with other

methods of faultline identification this can be a very subjective process (see Marsh, Ludtke,

Trautwein, & Morin, 2009; Petras & Masyn, in press; Vermunt & Magidson, 2002).

Unlike other methods of faultline identification LCCA is model-based, which means that

just as in structural equation modeling (SEM) a model is specified and fit to the observed data is

estimated. This allows LCCA to act in both a confirmatory and exploratory fashion. In an

exploratory analysis, general models can be estimated and the best fitting model selected. More

confirmatory approaches could involve specifying latent class means, variances, or other model

parameters and conducting likelihood ratio difference tests across nested models. Such an

approach would work toward avoiding the "dustbowl empiricism" associated with many

exploratory data analytic approaches.

LCCA is also probabilistic because class membership is unknown and each individual

has a probability of membership for each class. These posterior probabilities for each individual

can be estimated and are often used to assign individuals to the class with which they have the

highest probability of membership. Just as with other methods of *faultline identification* (e.g.,

Bezrukova, et al., 2009) there is uncertainty in this process and information regarding the

uncertainty of membership is lost in the process of assignment (Vermunt & Magidson, 2002).

However, to examine the justification for classifying individuals into their most likely class of

membership, statistics such as entropy may be used. The entropy statistic reflects the extent to which posterior probabilities of membership are high for one class and low for all other classes (DiStefano & Kamphaus, 2006). This means that the entropy statistic reflects the extent to which individuals are similar to one another in terms of their probabilities of class membership. For example, if two individuals have a probability of 1.0 of belonging to a given class, they have observed scores identical to the model-estimated scores for that class and, therefore, they would be similar. As such, the entropy statistic measures *faultline strength* based on the comparison of model-estimated subgroup parameters and individuals' data, a statistic that would be of interest to researchers desiring to know attribute alignment across subgroups identified by LCCA.

Importantly, the LCCA model can take many forms. The most common model is referred to as the local independence model (Vermunt & Magidson, 2002). As with other latent class models this can be thought of as a measurement model, again just as with various SEMs, where the covariance among the observed variables is explained by the latent class variable. In this model the covariance among observed variables is specified as zero, meaning the covariance among them is completely explained by the latent class variable (Croon, 2002). This fits with Lau and Murnighan's (1998) notion that covariance among observed variables indicates membership in latent subgroups that are not directly measured but serve to indicate the latent classes that define the faultlines in any group of interest.

However, and unlike other faultline methods, this model is easily extended to include complex structural components, such as categorical and continuous covariates that can be used to predict and explain class membership, as well as increase degrees of freedom to help model identification (Dayton & Macready, 2002; DiStefano & Kamphaus, 2006). These models can also accommodate latent factors. Such models may be referred to as structural equation mixture

models, mixture regression models, or finite mixture models. In these models, measurement and structural regression parameters can be estimated within each latent class (Wedel & DeSarbo, 2002). This suggests new avenues of investigation for faultline researchers interested in estimating relationships among variables while also estimating latent class membership.

Conveniently, LCCA allows for the integration of continuous and categorical variables with a number of underlying distributions 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. LCCA results are not adversely affected by the scale and variance of observed variables (see DiStefano & Kamphaus, 2006; Vermunt & Magidson, 2002), which, as noted earlier, has been discussed as an important issue in faultline research (Lau & Murnighan, 1998). Extant measures manage this by making continuous variables discontinuous (e.g., Shaw, 2004) or by transforming discontinuous variables (e.g., Bezrukova et al., 2009). Therefore, LCCA may be useful for faultline analysis because researchers need not weight or manually transform the variables. This means that faultline identification is unaffected by differences in variables' scaling. Moreover, it means a measure of faultline strength may be computed in the form of the entropy statistic without having to rescale observed variables.

Although LCCA can require more individuals than there are observed variables in some statistical packages, this is not a requirement and is facilitated with a full information maximum likelihood estimator (Enders, 2001; see similar thought in Hamaker, Dolan, & Molenaar, 2003). In traditional latent class modeling large samples are required when generalizing to a larger population, with additional individuals allowing for more stable class enumeration and unbiased estimates, i.e. these solutions are asymptotically correct. However, when such generalizations are

not desired it is possible to use LCCA in small groups (for example, see Barkema & Shvyrkov, 2007). This is important because in faultline research the observed groups are treated as the population of interest. No generalization to a larger population is desired.

In summary, LCCA is a useful technique for studying faultlines in large groups such as organizations and can be used in small groups such as those studied in traditional faultline research. It (a) is model-based and probabilistic, (b) can act as a measurement model for uncovering the latent faultlines that exist in groups, (c) is not sensitive to the scale of observed variables in both estimating the appropriate number of subgroups and providing a measure of faultline strength in the entropy statistic, (d) allows the inclusion of variables with a host of underlying distributions, (e) can be extended to include a wide variety of models, such as those with latent factors and discontinuous and continuous covariates with complex structures, and (f) can accommodate the analysis of large and small groups.

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 interactions as this indicates that the subgroups are socially-meaningful.

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 common 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. Gender and ethnicity variables reflect proportions with a lower-bound of 0.0 and an upper bound of 1.0.

These attributes are primarily surface-level attributes, which makes them likely candidates for creating social categories in an organization (Harrison, Price, Gavin, & Florey, 2002). In addition, research consistently suggests that such attributes are related to interaction.

For instance, relevant for this study, Zenger and Lawrence (1989) found that age and tenure were related to technical communication frequency.

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 indicate that they 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 his or her work-related contacts who fall within other subgroups. Work-related communication contacts are out-degree or asymmetric measures. 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 latent boundaries between subgroups, defined by positive associations

among individual attributes. 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 and consonant with recommendations for both quantitative and

qualitative LCCA assessment (Muthen, 2003), the attributes that seem to play the most

consistent role in subgroup differences are assessed qualitatively after the LCCA analysis..

Table 1 provides descriptive statistics of all individual-level variables, including means,

standard deviations and correlations.

Table 1 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. A

maximum likelihood estimator with standard errors robust to non-normality was used to account

for any non-normality in the observed variables. Additionally, all LCCA models were specified

as local independence models with all covariances among observed variables constrained to zero within each class. As noted above this makes our model similar to a measurement model, where covariance among observed variables, the basis of faultlines (Lau & Murnighan, 1998), is captured by the latent class variable.

We evaluated seven models, ranging from two to eight classes, 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 examine 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

Perhaps 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 functions of the average posterior probabilities of class membership across classes (B.O. Muthen & Muthen, 2000). If individuals have equal posterior probabilities of membership across all classes, they cannot be meaningfully assigned to a single class, indicating poor classification quality. Alternately, if individuals have a high probability of membership in a single class, while a low probability of membership in all other classes, they are easily assigned to a single class, which makes classification quality high. 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 (B. O. Muthen, 2003). Our theoretical interest was to identify structure in individuals' broad social frames of reference at work rather than just their workgroups. 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 from two- through 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 (for a similar method see Petras & Masyn, in press).

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 consistent with 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 structure 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 2). A univariate ANOVA across the five subgroups for each attribute shows that with two exceptions, the proportions of Black and Hispanic members, statistically significant F values suggesting differences in faultline strength across subgroups. The F values are measures of faultline strength. Given the significance of these global F tests, mean comparisons of subgroup pairs were performed to further indicate faultline strength. Subgroup pairs were first tested for equality of variances and mean comparison tests were adjusted accordingly.

Table 2 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 are more likely to be 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 identify should influence interaction among subgroup members. Our criterion-related assessment of validity is the extent to which subgroup membership is related to within-subgroup work-related communication frequency. The results in Table 3 support this assessment. Subjects consistently report more work-related communications in their own subgroup than in other subgroups, even after controlling for their own attributes. For instance, subjects' attributes explain 13% of the variation in their communications with Middle-Timers. Adding subjects' organizational subgroup increases explained variation to 30%. Middle-Timers communicate significantly more often with other Middle-Timers than do Old-Timers, fast-track men and high-level Old-Timers. The association between Asian Women Newcomers and Middle-Timer communication frequency is not significant. However, the pattern of higher within- than between-subgroup communication is consistent across all subgroups.

Table 3 About Here

It is possible that these results represent subjects who communicate only within their own subgroup and perhaps with one other subgroup. In order to get a better picture of the distribution of subgroup interactions, we examined the frequency of work-related communication contacts subjects reported across subgroups.[3] Middle-Timers appear the least isolated subgroup. Sixty-two percent of their reported work-related communication contacts fall in other subgroups. Thirty-two percent of their total contacts are with High Level Old-Timers and the remaining 30% are distributed in roughly equal numbers across the other three subgroups. Old-Timers are one of the two most isolated subgroups. Only 34% of their reported work-related communication contacts fall within other subgroups. Of their total contacts, 26% are with High Level Old-Timers. Old-

---

[3] These data are available from the first author.

Timers have few contacts who are Middle-Timers or Fast Track Men and none who are Asian Women Newcomers. The second most isolated subgroup is Asian Women Newcomers. Only forty-four percent of their work-related contacts fall within other subgroups. Of their total contacts, 35% are with Middle-Timers. The remaining nine percent are with Fast-Track Men or High Level Old-Timers. Asian Women Newcomers report no work-related contacts with Old-Timers.

These work-related communication results are consistent with the subgroup attributes reported in Table 3, showing that career level, age, organizational tenure, and education play important roles in these organizational 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.

One explanation for the strong within-subgroup communication results may be that subgroup members belong to the same formal work groups and are thus required to communicate with one another. We used same-supervisor to identify formal work groups and examined the extent to which individuals within each subgroup share supervisors. With the exception of one

group of six Asian Women Newcomers who work for the same supervisor and one group of five

fast-track men, no subjects share more than two supervisors with others in their subgroup. On

average, 71% share no supervisors with others in their subgroup. This suggests that although

common workgroups likely explain some of the observed within-subgroup communication

patterns, it is unlikely they dominate the results.

Discussion

This study explores the use of LCCA for examining organizational faultlines: attribute

alignments in a large group whose boundaries are defined by the attributes of others whom

individuals work with, talk to, hear about or observe. The faultline concept (Lau & Murnighan,

1998) 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 several 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 attribute alignments produce group-level phenomena. Existing faultline measures, such

as *Fau* and *FSL*, are designed for examining faultline strength and only *Fau* directly identifies

subgroups. Moreover, neither allows simultaneous analysis of attributes holding different scales and underlying distributional properties without manually transforming raw data.

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. It allows inclusion of attributes measured using nominal, categorical and continuous variables in the same analysis without additional categorization or transformation. LCCA also serves as a measurement model to investigate the latent faultlines that may exist in a group. In this respect it is model-based and probabilistic, allowing tests of models with varying numbers of subgroups to identify the model with the best fit, as well as allowing the computation of probabilities of membership for each individual for each subgroup. Finally, LCCA can be extended to other more complex specifications, allowing the inclusion of structural components.

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 results identify subjects whose organizational reference groups are most similar in composition.

The LCCA resulted in a five-class solution. The attributes of the subjects in each of the five subgroups differ statistically along each of the six attributes, with the exception of proportion Black and proportion Hispanic. A qualitative assessment of these differences consistent with company history suggest 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, appear to contribute more to subgroup identification than the nominal attributes, including gender and ethnicity. This is surprising given the tremendous attention the literature gives to gender and ethnicity as significant diversity attributes. The criterion-related validity of these subgroups was assessed using their reported work-related communication contacts. As expected, subjects' reported more within- than between-subgroup contacts.

We do not expect these faultlines to generalize to other organizations. Independent of demographic distributions, the meaning attached to aligned attributes is likely to differ. In the organization studied here, time seems to play a large role in informal social structure. This may result because average employees have had ample time to observe others and make sense of how things work. The salience of these understandings may have been enhanced by increasing competition in the organization's environment. The organization responded to this competition by hiring younger, more highly educated managers in higher-level positions than had been customary. As a result, the attributes of valued employees shifted, providing a visible contrast with the past and increasing the salience of differences. This suggests it may be important to consider how "history" influences attribute salience or creates new salient attributes. Moreover, it suggests that laboratory studies may overstate the importance of attributes such as gender and ethnicity because it may prove difficult to simulate "years of shared experience."

This last limitation suggests several questions. For instance, how does an organization's age influence its faultlines? Attribute alignments that are reinforced over many years may facilitate larger subgroups than those observed in a start-up. However, once subgroups get sufficiently large, they probably represent one of two extremes. Either they become representative of the organization's population or the correlations among attributes become so strong that faultline analysis is required on only a few attributes. This might result because the composition of organizations tends to become more homogeneous over time (Kanter, 1977a; Schneider, Goldstein, & Smith, 1995).

Future Theoretical Directions for LCCA

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 only groups of individuals.

Zyphur (2009) 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 influences performance. However, using LCCA one might ask whether interdependencies among performance and personality traits 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 many interaction terms.

At a different level of analysis, LCCA is appropriate for exploring the intermediate social structures that exist between individuals and organizations, a territory about which little is known. This study focused on the individuals in each subgroup, but individuals were not placed into subgroups because their attributes align. They were placed into subgroups because the attributes of their organizational reference groups align. Thus, instead of studying individuals who know similar others, we could study subgroups of similar others as they are perceived by organizational members. The individual-level study focuses on the individuals who perceive their organization differently. A group-level study would describe the informal structures these individuals are observing, where each subgroup represents a type of neighborhood. Focusing on the attributes of these subgroups might provide a new perspective on what emergent social structures look like in an organization.

This approach might be used to explore cultural identities in organizations, which result from groups "that are socioculturally distinct" (Ely & Thomas, 2001, p. 230). Such groups are often associated with members' demographic attributes, which tend to acquire distinct shared meaning regarding norms, values and expectations (p. 230). Ely and Thomas (2001) interviewed employees in three firms, one with 12 employees and the other two with around 110. They identified three distinct perspectives regarding cultural identity: integration-and-learning, access-and-legitimacy and discrimination-and-fairness. These perspectives were shared among workgroup members, but differed across workgroups within each organization. Although all three perspectives seemed to improve group functioning, only the integration-and-learning perspective produced long-term benefits. This inductive analysis was based on interviews.

However, adding LCCA would allow the authors to include a confirmatory analysis of whether the three perspectives do indeed differentiate workgroup's cultural identities. *Fau* would be less useful here because it is not model-based and could not be used to compare the results using different numbers of subgroups.

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

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster Analysis*. Beverly Hills, CA: Sage.

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.

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

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17*(5), 475-482.

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. *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, 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.

Croon, M. (2002). Ordering the classes. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 137-162). Cambridge, UK: Cambridge University Press.

Dayton, M. C., & Macready, G. B. (2002). Use of categorical and continuous covariates in latent class analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis*. Cambridge, UK: Cambridge University Press.

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.

Ely, R. J., & Thomas, D. A. (2001). Cultural diversity at work: The effects of diversity perspectives on work group processes and outcomes. *Administrative Science Quarterly, 46*(2), 229-273.

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

Fiske, S. T., & Taylor, S. E. (1991). *Social Cognition* (2nd ed.). New York: McGraw-Hill.

Goethals, G. R., & Darley, J. (1977). Social comparison theory: An attributional approach. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 259-278). Washington, D.C.: Halsted/Wileuy.

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.

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, teams, and task performance: Changing effects of surface- and deep-level diversity on group functioning. *Academy of Management Journal, 45*(5), 1029-1045.

Hogg, M. A., & Terry, D. J. (2000). Social identity and self-categorization processes in organizational contexts. *Academy of Management Review, 25*, 121-140.

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.

Kanter, R. M. (1977a). *Men and Women of the Corporation*. New York: Basic Books.

Kanter, R. M. (1977b). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology, 82*, 965-990.

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.

Lawrence, B. S., & Tolbert, P. S. (2007). Organizational demography and careers: Structure, norms and outcomes In M. Peiperl & H. Gunz (Eds.), *Handbook of Career Studies* (pp. 399-421). Thousand Oaks, CA: Sage Publications.

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.

Linton, R. (1936). *The Study of Man*. New York: D. Appleton-Century.

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

Marsh, H. W., Ludtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling, 16*, 191-225.

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

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

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.

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.

Petras, H., & Masyn, K. (in press). General growth mixture analysis with antecedents and consequences of change. In A. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology*.

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.

Polzer, J. T., Milton, L. P., & Swann, W. B. (2002). Capitalizing on diversity: Interpersonal congruence in small work groups. *Administrative Science Quarterly, 47*, 296-324.

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.

Ridgeway, C. (1991). The social construction of status value: Gender and other nominal characteristics. *Social Forces, 70*(2), 367-386.

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.

Schneider, B., Goldstein, H. W., & Smith, D. B. (1995). The ASA framework: An update. *Personnel Psychology, 48*, 747-773.

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.

Trice, H. M., & Morand, D. (1991). Organizational subcultures and countercultures. In G. Miller & C. K. Warriner (Eds.), *Studies in Organizational Sociology : Essays in Honor of Charles K. Warriner*. Greenwich, CT: JAI Press.

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

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 89-106). Cambridge, UK: Cambridge University Press.

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.

Wedel, M., & DeSarbo, W. S. (2002). Mixture regression models. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 366-382). Cambridge, UK: Cambridge University Press.

Wheeler, L. (1991). A brief history of social comparison theory. In J. M. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 3-21). Hillsdale, NJ: Lawrence Erlbaum.

Zanutto, E. L., Bezrukova, K., & Jehn, K. A. (in press). Revisiting faultline conceptualization: Measuring faultline strength and distance. *Quality & Quantity*.

Zenger, T. R., & Lawrence, B. S. (1989). Organizational demography: The differential effects of age and tenure distributions on technical communication. *Academy of Management Journal, 32*, 353-376.

Zyphur, M. (2009). When mindsets collide: Switching analytical mindsets ot advance organizational science. *Academy of Management Review, 34*(4), 677-688.
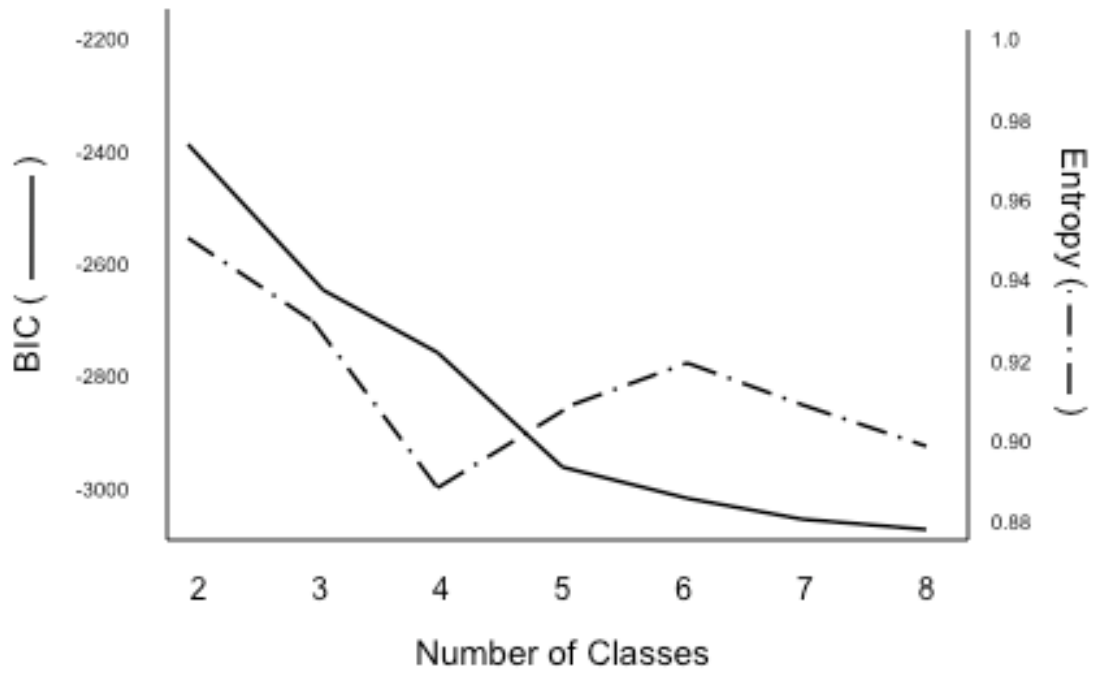
Figure 1. **BIC and entropy values for two- through eight-class models. Lower values indicate better fit.**

TABLE 1
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
Work-Related												
19. 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 1 (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	
Work-Related 19. 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 2
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 3
Regression of Number of Others With Whom Subject Communicates About Work on
Individual Attributes and Organizational Subgroup

| | Number of Others With Whom Subject Communicates About Work Who Are: | | | | | | | | | |
| | Middle-Timers | | Old-Timers | | Fast Track Men | | High Level Old-Timers | | Asian Women Newcomers | |
	B	se	B	se	B	se	B	se	B	se
Step 1: Individual's Attributes										
Gender[a]	0.73 ***	0.17	-1.36 ***	0.27	-0.14	0.11	0.92 ***	0.21	-0.04	0.14
Black	0.05	0.27	-0.22	0.42	-0.13	0.17	-0.28	0.33	0.32	0.22
Hispanic	0.35	0.22	-0.52	0.33	-0.12	0.14	0.57 *	0.27	0.07	0.17
Asian	-0.05	0.25	-1.13 **	0.39	-0.11	0.16	-0.50	0.31	1.59 ***	0.20
Age	-0.02	0.02	-0.02	0.02	-0.02 *	0.01	0.03	0.02	0.00	0.01
Organizational Tenure	-0.02	0.01	0.06 **	0.02	-0.01	0.01	0.04 *	0.02	-0.04 **	0.01
Education	0.07	0.09	-0.58 ***	0.14	0.06	0.05	-0.04	0.11	0.13 †	0.07
Career Level	0.06 †	0.03	-0.23 ***	0.05	0.06 ***	0.02	0.13 ***	0.04	-0.03	0.02
F	6.76 ***		21.72 ***		6.96 ***		10.16 ***		20.56 ***	
R^2	0.13		0.33		0.14		0.19		0.32	
Step 2: Organizational Subgroup										
Middle-Timers	*b*		-3.41 ***	0.34	-1.08 ***	0.18	-1.06 ***	0.31	-2.65 ***	0.24
Old-Timers	-1.98 ***	-0.17	*b*		-1.88 ***	0.20	-1.13 ***	0.26	-3.05 ***	0.27
Fast Track Men	-1.16 ***	0.29	-3.85 ***	0.49	*b*		-2.07 ***	0.44	-3.24 ***	0.26
High Level Old-Timers	-0.99 ***	0.24	-2.14 ***	0.30	-1.45 ***	0.21	*b*		-2.94 ***	0.28
Asian Women Newcomers	-0.43	0.33	-3.97 ***	0.55	-1.66 ***	0.22	-2.41 ***	0.49	*b*	
F	12.65 ***		29.93 ***		15.01 ***		11.06 ***		36.00 ***	
R^2	0.30		0.50		0.34		0.28		0.56	
ΔR^2	0.17 ***		0.18 ***		0.20 ***		0.09 ***		0.24 ***	

[a] Minorities, including women and ethnic groups, are coded = 1.

[b] Comparison category.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$.