Team Faultline Measures: Rescaling the Weights of Diversity Attributes

Keivan Bahmani
Electrical and Computer
Engineering
Clarkson University
bahmank@clarkson.edu

Zhaleh Semnani -Azad
Reh School of Business,
Electrical and Computer
Engineering
Clarkson University
zsemnani@clarkson.edu

Katia Sycara
Robotics Institute
Carnegie Mellon
University
katia@cs.cmu.edu

Michael Lewis
School of Information
Sciences
University of Pittsburgh
ml@sis.pitt.edu

Abstract

Faultline, or subgroup formation based on the alignment of diversity attributes, can cause conflicts and low coordination in diverse teams. While researchers understand the importance of faultlines in team process and negotiations, current computational faultline measures are highly vulnerable to subjective weight assignment of diversity attributes. Therefore, there is limited understanding of which diversity attributes have more impact on faultline formation. In this paper we report 1) a pilot study illustrating the susceptibility of the current faultline measures to subjective evaluations, and 2) an online study illustrating how people's surface (e.g. age, gender, race) and deep (e.g. personality, cultural norms) level diversity attributes impact their preference and selection of team members, as a proxy of faultline formation. We find while various surface and deep-level attributes predict selection of members, most of these attributes are highly correlated with members' age, suggesting the importance of this attribute. We discuss future directions for faultline measures with objective rescaling of diversity weights.

1. Introduction

Today's ever-growing globalization trend has encouraged many organizations to construct and rely on diverse teams to compete in a global market [1]. Diverse teams include a group of people with various *surface-* (visible demographic characteristics such as age, gender, and ethnicity) and *deep-* (invisible elements such as norms and values) level diversity attributes, committed to a common goal [2]. Diverse teams are essential to organizational innovation, creativity and productivity [3]. For instance, racial diversity in a team has been associated with higher team performance [4]. While diversity in organizations yields many benefits, it also gives rise to major problems such as team conflict, lower communication effectiveness, and lower collaboration and unity [5]. For instance, diversity can increase relationship conflict, or deep-rooted personal differences in values, beliefs, and personalities [6]. This type of conflict is more likely to hinder communication and team performance [5].

Faultline is a major contributing factor to relational conflict in diverse teams. *Faultlines* are hypothetical, dividing lines based on the alignment of surface- and deep-level diversity attributes that result in the formation of within team subgroups [7]. For example, a four-member team composed of varied age range and education background can yield two subgroups of younger versus older members, with age becoming the most pronounced attribute in the faultline formation. Within team, subgroups can negatively affect team process and outcome through in-group bias [7], [8]. This is when a subgroup has favoritism toward its members, along with negative perception toward the out-group members, i.e. the other subgroup in the team [9]. The stronger the in-group/out-group distinction, the higher the relationship conflict, and the lower the team unity, collaboration, communication, and performance [7], [8]

Faultlines are important to understand and predict, yet their relationship with conflict and performance is not always constant. For instance, depending on the level and type of diversity attributes, teams can have several potential faultlines that are not yet perceived or noticed, i.e. *dormant faultlines* [7]. Once perceived, i.e. *activated faultlines,* their impact on conflict and outcome become more detrimental, because of reduced cross-subgroup communication and collaboration [10]. Moreover, faultlines tend to have a curvilinear relationship with conflict, morale, and performance [11]. Highest level of conflict arises in teams with very high diversity, i.e. no faultline, or even alignment of diversity attributes with homogenous subgroups, i.e. strong faultlines. Medium

HICSS

level faultlines have weaker relationships with conflicts. Consequently, an accurate measure is required to determine the level or strength of faultline by taking into account 1) the type of diversity attributes and 2) the weight and impact of those attributes on diversity alignment and subgroup formation. Such a measure can help researchers predict and manage conflicts and team outcomes.

To date, researchers have developed several algorithms to measure faultlines. The three most frequently used algorithms include: *Fau* [11], *Faultline Distance* [12], and *Average Silhouette Width (ASW)* [13] (see Table 1). All three algorithms rely on rescaling weights to combine various diversity attributes into one common value to calculate the final faultline strength. Accordingly, these algorithms are dependent on the subjective evaluation and input of diversity weights. This means such measures are more prone to human bias. Indeed, researchers do recognize the limitations of such measures and have stressed the importance of these rescaling weights [11], [14]. Yet, there is a dearth of work on these effects or a systematic process to evaluate them. Furthermore, there is limited research illustrating a more objective evaluation of the weights and the impact of various diversity attributes on faultline formation.

To extend prior work on faultline measures, in this paper we first shed light on the effects and inconsistencies associated with the rescaling of weights employed in Fau, Faultline Distance and ASW. In a pilot study, we include a subset of our previously collected data on diverse teams engaging in a negotiation simulation. We extract faultline strength using the three faultline measures and rescale the weights of diversity attributes. As expected, the subjective evaluation of the weights have a profound effect on the relationship strength and direction between diversity attributes and the faultline value.

Second, we expand on the faultline and diversity literature by extracting objective weights of the diversity attributes and their impact on faultline formation. In an online, inductive, hypothetical team study with human subjects from Amazon's Mechanical Turk (M-Turk), we examine the impact of surface and deep-level diversity attributes of a person on their perception and preference toward other individuals varying on diversity attributes. This paradigm is used to capture in-group preference based on diversity characteristics as a proxy for faultline and subgroup formation. We report various relationships between surface and deep-level attributes, and the potential weights associated with these factors in forming faultlines.

2. Faultline Measures

In the past few decades, researchers have developed various faultline measures, most calculating faultline strength derived from potential dormant faultlines [7], [11]–[14]. *Faultline strength* captures the strength of members' attachment to the group based on potential homogeneity of the subgroup [12]. Faultline strength depends on three factors: number of demographic attributes apparent to the team members, alignment of the attributes, and the number of resulting homogenous subgroups. Prior measures organized faultline strength into five categories of no faultline, very weak, weak, strong, and very strong faultline [7]. The higher the strength of faultline, the more attributes are aligned in the team, resulting in higher homogeneity within the subgroups.

The most commonly used faultline measure is Fau [11]. Fau exhaustively examines all the possible subgroups, calculates the total variation in overall group characteristics (Fau_g), and chooses the subgroup with the highest ratio of group sum of squares over total sum of squares (see Table 1). For instance, in a team where members vary on two diversity attributes of Age and Education, using Fau the researcher can align the change in 10 years of age with a level of change in education, e.g. C (0.1, 1). This means that a change in education level contributes to as much strength as 10 years in age. Fau then implements two dummy variables for education, multiplies those values by 1/square root of 2, and divides the age values by 10. The Fau measure can take any value from *0* (no faultline) to *1* (very strong faultline). The positive aspect of Fau lies in the process of subgroup selection and that it paved way for other measures. Yet, it is limited in the number of subgroups it considers for the calculation. Thus, the valuation reflects an underestimated faultline strength, particularly in large teams with more than two possible subgroups.

To improve some of the limitations of Fau, Bezrukova and colleagues [12] proposed another algorithm that calculates distance as an additional index of faultline. *Faultline Distance* denotes the differences between aligned subsets or subgroups within a team, generated from faultlines [12]. Similar to strength, distance can change the impact of faultline on team processes and outcomes [12]. Large differences or gaps among subgroups can significantly reduce cross-subgroup interaction, communication, and cohesion. Yet, Faultline Distance yields additional information beyond strength. For example, if there are two teams with clear subgroups based on gender, ethnicity, and age, Fau can treat all these dimensions equally (i.e. all alignments will be equated to 1),

Table 1. Summary of the most frequently used faultline measures.

Name	Description	Formula
Thatcher's Fau [11]	The algorithm finds a two-subgroup configuration associated with largest ratio of between group variance over the total group variance. Note that this approach only works with two subgroups.	$Fau_g = \left[\dfrac{\sum_k \sum_j n_k^g (\bar{x}_{.jk} - \bar{x}_{.j.})^2}{\sum_k \sum_j \sum_i (x_{ijk} - \bar{x}_{.j.})^2} \right]$
Faultline Distance [12]	Considers an extra measure called *distance* in addition to FAU strength. As a result, this approach also only works with two subgroups.	$D_e(X,Y) = \sqrt{\sum_i (x_i - y_i)^2}$
Average Silhouette Width (ASW) faultline [13]	Uses hierarchical cluster analysis techniques to detect the subgroups leading to the strongest faultline.	$S(i) = \dfrac{b_i - a_i}{\max(a_i, b_i)}$

While Faultline Distance can recognize additional differences along these dimensions (e.g. making a distinction between age gap of 20-50 versus 20-30 in the two teams). Faultline Distance computes the product of the strength of faultlines derived from the Fau and the Euclidian distance between the means of each attribute in subgroups as a faultline measure. Accordingly, Faultline Distance captures another dimension of faultline, which is not recognized in Fau.

Both the Fau and Faultline Distance exhaustively look for all possible subgroup combinations. As a result, the number of calculations required for these algorithms to extract all the possible subgroups of an *n* member team is equal to the Bell number (B_n). B_n is an extremely fast growing number, which drastically limits the usability of Fau and subsequently Faultline distance algorithms for large teams. Thatcher et al. addressed this issue by limiting the number of subgroups in the Fau measure to only two subgroups [11]. Yet, this has become a major limitation in both Fau and Faultline Distance measures, as they underestimate faultline strength in teams with more than two homogeneous subgroups [14].

Extending FAU and Faultline Distance, Meyer and colleagues proposed the *Average Silhouette Width (ASW)* faultline measure [13]. ASU categorizes team members into subgroups with the maximum internal subgroup homogeneity and between subgroup heterogeneity. The ASW algorithm has two main steps. First, the algorithm employs hierarchical cluster analysis (agglomerative cluster algorithms), more specifically Ward and average link strategy, to find the initial set of subgroups. Second, the algorithm permutes and re-arranges team member composition to extract the maximum ASW, which in turn identifies the subgroup split with the strongest faultlines (Table 1). Accordingly, ASW addresses the limitations of Fau and Faultline Distance by calculating a faultline value for teams with more than two subgroups. Moreover, prior work on group faultline confirms that *configurational properties* of teams, i.e. the number and variation in the size of subgroups, can affect team outcome and relational conflict [15]. As a result, ASW can be used in bigger teams with higher number of potential subgroups to generate a more accurate measure for faultlines [14]

The aforementioned algorithms use a similar rescaling mechanism to combine the values from different attributes into one distance measure to quantify team faultline. This process consists of two steps. First, c dummy variables are employed to represent the distance between *c* categories (categorical values) that are equal, with c-1 denoting the uneven distance between categories [11]. Second, the *c* dummy variables and the remaining continuous variables are rescaled and combined into one distance measure [11], [14]. Each algorithm calculates the product of each attribute by its associated weight. For instance, Thatcher et al. [11] combined 10 years of age and a difference in gender or race equally. This combination respectively leads to the weights c (0.1, 1, 1) for age, gender and ethnicity. Thatcher et al. suggest that their rescaling of weights in this example is reasonable. They further argue that researchers can treat such weights and rescaling of the diversity attributes subjectively, especially in the absence of information on these weights. Yet, they also call for future research to examine the effects of these rescaling factors.

In our work, we examine the impact of weights on the faultline valuation derived from Fau, Faultline Distance and ASW. We speculate that the rescaling of weights depend on the underlying *surface* and *deep-level* diversity attributes of the team. For instance, all three algorithms calculate the same faultline strength for the same team composition in two different

countries such as USA and India without considering the vast underlying demographic and cultural differences. Accordingly, these algorithms may not consider additional underlying factors, particularly deep-level diversity attributes, and their impact on the activation of faultlines.

There may also be variation in weights associated with surface-level diversity attributes, and these can differ across teams. Social identity theory suggests people have a tendency to categorize themselves and others into groups based on shared social categories such as age, gender and ethnicity [9], [15]. Such categorizations can generate identity-based subgroups, with individuals perceiving the in-group members sharing a common identity as themselves [15]. Yet, when people belong to multiple social categories, individuals classify these people either on a single dominant category (e.g. age) [16], or via the additive combinations or intersections of these categories [17]. This theoretical framework, along with prior empirical studies suggest that there are differential weights associated with surface-level diversity in subgroup formation, and accordingly faultline activation. For instance, in a team where majority of its members categorize others based on age, then the likelihood of faultlines activated based on the alignment of age is much higher compared to other attributes such as gender and race. Accordingly, age should be given a higher weight than the other attributes.

Overall, it is important for researchers to 1) check the impact of rescaling and sampling of weights associated with the three faultline measures, and 2) define the weights in a more objective manner to better predict faultline formation and activation. We test the effect of the rescaled weights on the faultline calculations in a pilot study. We then conduct another study with general working population to derive a more objective understanding of weights associated with surface and deep-level diversity attributes on in-group preference and potentially subgroup formation, as an index of faultline activation.

3. Pilot Study

The purpose of this study is to examine how rescaling of weights associated with surface-level diversity attributes impact faultline strength. The valuation of faultline is associated with the three prominent faultline algorithms of Fau, Faultline Distance and AWS[11]–[13] .We initially start with equal weights of the three commonly used attributes: gender, age, and ethnicity [11]. We then vary the weights associated with each attribute, while controlling for the other features. We report the final

faultline valuation based on the sample of weights. We also examine the extent to which the rescaled weights and the derived faultline strength can predict faultline activation, or perceived faultline, as reported by team members in this negotiation study.

3.1 Method

Participants. Participants were 97 undergraduate management students (52.6 % female, Mean age= 21.48, S.D. = 1.54) from two North-American Universities. Most participants were Caucasians (68.1%). We also had East Asian (16.5%), Middle Eastern (7.2%), African American (4.1%), Latin American (2.1%) and South-Asian (2.1%) participants.

Procedure. Participants engaged in a supply-chain management dispute negotiation [18]. Participants were placed in a negotiation team of four members varying in gender, ethnicity and age. A week before the negotiation exercise, participants read about their roles and prepared for their first, intra-group interaction about planning and implementation of strategies for the negotiation. The team interactions were face to face and lasted around two hours. We provided participants with a survey after this stage to measure faultline activation, or the extent to which they noticed they team split into smaller subgroups. We used the four-item activated group faultline measure [10] to capture active faultlines (α = .99). This measure captures the extent to which individuals notice subgroup formation in their teams based on diversity attributes.

3.2 Results and Discussion

We calculated faultline strength based on the team variation of diversity attributes of age, gender and ethnicity. We carried out these calculations using the existing faultline algorithms [19]. We then regressed the faultline strength outputs associated with the rescaled weights to the self-reported faultline activation. We conducted linear, hierarchical regression analyses to derive these correlational outputs. We wanted to examine 1) how faultline strength varies depending on the sampling of weights, and, 2) how faultline strength, associated with dormant faultlines, predicts active and perceived faultlines in the teams.

Table 2 includes all our outputs of the rescaled weights associated with faultline strength and regression outputs associated with faultline activation. We can clearly observe that the variation in the weights has a considerable effect on the strength of faultlines calculated with all three algorithms.

Table 2. Regression Analysis for FAU,
Faultline Distance and ASW

Algorithm	Weights (Gender/Age/Ethnicity)	β Coefficient	p
Fau	C (1,1,1)	-0.19	0.74
	C(1,0.1,1)	-0.115	0.46
	C (1,0.7,0.7)	-0.44	0.62
	C (1,0.4,0.4)	**-0.21**	**0.04**
	C (1,0.1,0.1)	0.35	0.70
	C(0.7,1,0.7)	-1.00	0.31
	C (0.4,1,0.4)	-1.70	0.06
	C (0.1,1,0.1)	**2.68**	**0.01**
	C (0.7,0.7,1)	2.20	0.21
	C (0.4,0.4,1)	-2.40	0.08
	C (0.1,0.1,1)	-0.02	0.98
Faultline Distance	C (1,1,1)	-0.19	0.74
	C (1,0.1,1)	-0.02	0.88
	C (1,0.7,0.7)	0.90	0.24
	C (1,0.4,0.4)	**-1.19**	**0.03**
	C (1,0.1,0.1)	0.24	0.73
	C (0.7,1,0.7)	0.355	0.78
	C (0.4,1,0.4)	0.640	0.62
	C (0.1,1,0.1)	**-1.00**	**0.02**
	C (0.7,0.7,1)	0.03	0.98
	C (0.4,0.4,1)	-0.66	0.44
	C (0.1,0.1,1)	0.80	0.53
ASW	**C (1,1,1)**	**-0.22**	**0.03**
	C (1,0.1,1)	-0.06	0.63
	C (1,0.7,0.7)	0.38	0.61
	C (1,0.4,0.4)	0.23	0.84
	C (1,0.1,0.1)	-0.72	0.47
	C (0.7,1,0.7)	-0.12	0.89
	C (0.4,1,0.4)	**1.98**	**0.03**
	C (0.1,1,0.1)	**-2.28**	**0.01**
	C (0.7,0.7,1)	-1.37	0.15
	C (0.4,0.4,1)	1.15	0.47
	C (0.1,0.1,1)	0.05	0.96

Furthermore, the strength of faultlines have different directions of relationships with activated faultlines. Across Fau and Faultline distance calculations, C (1, 0.4, 0.4) and C (0.1, 1, 0.1) associated with gender, age, and ethnicity, yielded significant correlations with faultline activations. Based on these effects we see that gender and age had more weight in predicting active faultlines. Within the ASW calculation, C (1, 1, 1), C (0.4, 1, 0.4), and C (0.1, 1, 0.1) were significantly related to faultline activation. Across these combinations, age, had more weight in the faultline strength and its relationship with activation.

A major limitation we observe is the direction of these relationships. For instance the rescaled weights of C (0.1, 1, 0.1) calculated with Fau (β = 2.68, SE = 3.84, t = 2.81, p =.01) and C (0.4, 1, 0.4) calculated

with ASW (β = 1.98, SE = 4.23, t = 2.21, p =.03) were the only combinations that had significant and positive relationships with faultline activation. Across these two combinations, age is weighted more heavily than gender and ethnicity. While age is also weighted more in the C (0.1, 1, 0.1) by Faultline Distance (β = -1.00, SE = .49, t = -2.21, p =.02) and ASW (β = -2.28, SE = 3.37, t = -3.00, p =.01), their relationships with faultline activation were negative.

The findings from this pilot study illustrate that the most commonly employed combination of weights c (0.1, 1, 1) for age, gender and ethnicity, heavily used in prior work [11], is not always ideal. In fact, in this pilot study we did not observe any relationship with faultline activation when implementing this set of weights. Instead, we found other set of weights yielding significant correlations with faultline activation, with age having the most weight over the other attributes. However, in some instances these weights have negative relationships with activation, suggesting that the stronger the faultlines, the lower chances of them being perceived by the team members. In other instances, this relationship is positive. Indicating that higher faultline strength positively predicts the active perception of faultlines. Accordingly, there is a need to understand the weight of attributes in a more objective manner.

4. Current Study

In our pilot study, we found that 1) the recommended weights of attributes reported by prior experiments [11] yield inconsistencies in faultline strengths, 2) the rescaled weights show different strengths and patterns of relationships with faultline activation, and 3) from the surface-level diversity, age seems to be an important factor contributing to faultline activation. We also see the need for 1) developing more objective measures to calculate the weights associated with diversity attributes, and 2) examining how deep-level attributes interact with surface-level diversity features in subgroup formation and faultline activation.

In the current study, we seek to address some of the inconsistencies observed in the surface-level diversity weights. We also examine which deep-level diversity attributes have a more important contribution to faultline formation and activation. In this study, we examine people's perception of other individuals who may become their teammates. We gathered participants' surface and deep-level attributes and provided a description of other people's surface-level attributes. We use such perceptions to understand in-group preference, as a proxy for subgroup formation and faultline activation. We examine the diversity

attributes and their relationship in selecting others. We also examine which diversity attributes contribute more to team member selection and preference. Based on the pilot study, we predict that from the surface-level attributes, age will have the most impact in predicting team member selection and preference. For deep-level attributes we examine personality [20] and cultural norms [21], [22], since these are most commonly examined in team and faultline research [11]. Accordingly, we shed light on the weights and importance of these attributes in subgroup formation.

4.1 Participants

We recruited 269 participants (39.8% female, Mean age= 35.23, S.D. = 11.11) from Amazon's Mechanical Turk (M-Turk) for a 15-minute online survey. Most participants were between 25 and 34 years of age (48.7 %), followed by 35 to 44 years (20.8%), 45 and older (19%) and 18 to 24 years (11.5%). Participants were South-Asian (50.9%), North-American (40.1%), European (3%), East Asian (3.0%), African (1.5%), South-American (1.1%) and Middle Eastern (0.4). Most participants obtained graduate/professional degrees (48.1%), bachelors (36.5%) and lastly high school diploma (15.4%).

4.2 Task

This online study captured participants' preference and perception of potential team members. Participants were told that they would be joining four other MTurk participants (i.e. four-member team) to engage in a problem-solving task. We mentioned that we would be gathering demographic data from all MTurk participants, including themselves, to find eight potential MTurkers for their team. We then asked them to rate the eight members presented and select which of the four they would prefer to work with on a problem-solving task.

We asked all participants to provide surface (age, gender, ethnicity, and education background) and deep-level (tight-loose, honor, face, and dignity cultural norm, and conscientiousness) diversity information. Accordingly, we captured features from both surface and deep-level diversity for all participants to examine the weight of these attributes on team member preference and selection.

After gathering participant diversity features, we presented participants with four surface-level diversity attributes of gender, ethnicity, education and age (see Table 3). We told participants to select which diversity attributes are important to them when working with teammates. We asked participants to rank this list of attributes on an 11-point ranking metric (1-11 points),

since the four attributes consisted of 11 dimensions in total (e.g. male, female, North-American, etc…). We indicated that we would consider their diversity information as well as their ranking of the attributes to present them with eight MTurkers for potential team members.

We then provided the participants with a list of eight potential team members and asked participants to select four members to join them in a team. The eight potential members shown to the participants corresponded to all of the possible combinations of the surface-level attributes in Table 3. However, we controlled the representation of ethnicity. Participants were randomly assigned to three conditions and one of the three sets of eight MTurkers: 1) North-American Mturkers varying on age, gender and education, 2) South-Asian MTurkers varying on age and gender, 3) four North-American and four South-Asian MTurkers varying on age and gender.

Table 3. List of attributes presented to participants

Diversity Attributes of MTurk Team Members

Gender	Male
	Female
Ethnicity	South-Asian
	North-American
Education	High school diploma
	Undergraduate degree
	Graduate degree
Age	18-24
	25-34
	35-44
	45-54

4.3 Measures

Table 4. contains a summary of the adopted measures and the corresponding alpha value associated with the measure's reliability. We measured surface-level diversity via four demographic attributes, of age, gender, ethnicity and educational background. We measured deep-level diversity via cultural norms associated with honor, face and dignity using the fifteen-item measure from prior literature [21]. We also measured tight-loose cultural norms using the six-item tightness-looseness scale by Gelfand and colleagues [22]. These measures shed light on people's perception of the strength of their social norms and tolerance for deviance. Higher score indicates higher endorsement of tight cultural norms. We used conscientiousness, as another measure of deep-level diversity. Conscientiousness is one of the big five personality attributes [20]. To measure conscientiousness we adopted the 10- item

Table 4. Summary of Measures

Measures	Sample Item	Alpha
Tight/Loose Norms	There are many social norms that people are supposed to abide.	0.75
Conscientiousness	I am always prepared.	0.75
Face	People should be very humble to maintain good relationships.	0.88
Honor	Are concerned about the reputation of their families.	0.76
Dignity	People should stand up for what they believe in even when others disagree.	0.85

conscientiousness measure by Goldberg and colleagues [23]

5. Results

We conducted a hierarchal regression analyses to examine the relationship between the surface and deep-level diversity attributes of participants (i.e. agents) with their preference and selection of team members, and how the selection was related to the members' surface-level diversity (see Table 5).

5.1 Agents' Surface-Level Attributes

We found interesting relationships between participant/agents' surface-level characteristics and their preference of members based on member surface-level diversity attributes (Table 5). We found a positive relationship between agents' age and members' age indicating that older agents preferred working with older members. Interestingly, older agents preferred to work with male members more than female members. We found that female agents

Table 5. Relationship between surface and deep-level attributes

Diversity	Agents' attribute	Member Attributes	β Coefficients	P value
Surface-level	Age	Male	.15	.01
		Female	.11	.06
		18 to 24 years old	-.39	<.01
		25 to 34 years old	-.36	<.01
		35 to 44 years old	.12	.04
		45 to 54 years old	0.49	<.01
	Gender	Male	-.15	.01
		Female	.33	<.01
		18 to 24 years old	-.11	.05
	Ethnicity	Male	.25	<.01
		18 to 24 years old	.14	.07
		35 to 44 years old	-.26	< .01
		45 to 54 years old	-.16	.02
		South Asian	.44	<.01
	Education Background	18 to 24 years old	-.23	<.01
		Graduate	.50	< .01
		High school Diploma	-.28	< .01
		Undergraduate degree	-.17	.01
Deep-Level	Tight and Loose Cultural norms	Female	.12	0.6
		25 to 34 years old	-.14	.03
		45 to 54 years old	-.14	.01
		South Asian	.15	.01
	Norm of Honor Culture	35 to 44 years old	.23	<.01
		45 to 54 years old	.12	.06
	Norm of Face Culture	45 to 54 years old	.14	.03
		North American	-.24	<.01
	Norm of Dignity Culture	45 to 54 years old	.12	.02
		Undergraduate degree	-.12	.05
	Norm of Conscientiousness	Female	-.14	.02
		45 to 54 years old	-.12	.02

preferred female members and older members. In contrast, male agents preferred working with younger members who are males. South Asian agents preferred working with other South Asian members, as well as younger male members. Highly educated agents also preferred working with members who are highly educated.

5.2 Agents' Deep-Level Attributes

We found interesting relationships between agents' deep-level diversity attributes and their selection of team members, based on members' surface-level diversity features (Table 5). We found a positive relationship between agents' endorsement of tight norms and preference for female and South Asian members. Agents endorsing honor, face and dignity norms were more likely to prefer working with older members. In addition, members endorsing face norms were less likely to select North American members. Members endorsing dignity norms were less likely to select members with an undergraduate degree. Interestingly, agents who scored high on conscientiousness were less likely to prefer female and older members.

6. Discussion and Future Direction

The results suggest that agent's both surface and deep-level attributes have an important effect on their perception of other team members' surface-level attributes. More specifically, the prominent effect of age in determining the strength of dormant faultlines and the fitness of faultline measures.

Our pilot study reveals that age has more weight, in determining the goodness of faultline measures in predicting perceived faultlines. Additionally, our main study confirms our speculation about the relationship between agent's surface and deep-level attributes and perception of other team members' surface-level attributes. In accordance with the result of our pilot study, our main study reveals significant relationship between agents' age and their perception of other team members' age and gender. Furthermore, the result also indicates the relationship between the endorsement of honor, face and dignity norms and perception of other team members' age. As a result, we believe the endorsement of honor, face and dignity norms has an influential effect on faultline formation and activation based on the alignment of surface-level attributes.

Our work confirms the influence of deep-level attributes on Fau, Faultline Distance and ASW faultline measures through the rescaling weights. In future work we aim to fully investigate this relationship, extend our study to more deep-level dimensions and accordingly propose a model that allow us to capture all the nuances in the relationship between people's surface and deep-level attributes.

Lastly, we aim to develop a novel faultline measure that objectively defines the rescaling weights based on the composition of the surface and deep-level attributes of each individual team by employing our comprehensive model.

7. Acknowledgment

8. References

[1] N. S. Wright and G. P. Drewery, "Forming cohesion in culturally heterogeneous teams: Differences in Japanese, Pacific Islander and Anglo experiences," *Cross Cult. Manag. Int. J.*, vol. 13, no. 1, pp. 43–53, 2006.

[2] N. DiTomaso, C. Post, and R. Parks-Yancy, "Workforce diversity and inequality: Power, status, and numbers," *Annu Rev Sociol*, vol. 33, pp. 473–501, 2007.

[3] S. K. Horwitz and I. B. Horwitz, "The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography," *J. Manag.*, vol. 33, no. 6, pp. 987–1015, Dec. 2007.

[4] O. Richard, A. McMillan, K. Chadwick, and S. Dwyer, "Employing an innovation strategy in racially diverse workforces: Effects on firm performance," *Group Organ. Manag.*, vol. 28, no. 1, pp. 107–126, 2003.

[5] K. A. Jehn, G. B. Northcraft, and M. A. Neale, "Why differences make a difference: A field study of diversity, conflict and performance in workgroups," *Adm. Sci. Q.*, vol. 44, no. 4, pp. 741–763, 1999.

[6] J. S. Chun and J. N. Choi, "Members' needs, intragroup conflict, and group performance.," *J. Appl. Psychol.*, vol. 99, no. 3, p. 437, 2014.

[7] D. C. Lau and J. K. Murnighan, "Demographic Diversity and Faultlines: The Compositional Dynamics of Organizational Groups," *Acad. Manage. Rev.*, vol. 23, no. 2, pp. 325–340, Apr. 1998.

[8] M. A. Hogg, J. C. Turner, and B. Davidson, "Polarized norms and social frames of reference: A test of the self-categorization theory of group polarization," *Basic Appl. Soc. Psychol.*, vol. 11, no. 1, pp. 77–100, 1990.

[9] H. E. Tajfel, *Differentiation between social groups: Studies in the social psychology of intergroup relations.* Academic Press, 1978.

[10] K. A. Jehn and K. Bezrukova, "The faultline activation process and the effects of activated faultlines on coalition formation, conflict, and group outcomes," *Organ. Behav. Hum. Decis. Process.*, vol. 112, no. 1, pp. 24–42, May 2010.

[11] S. M. Thatcher, K. A. Jehn, and E. Zanutto, "Cracks in diversity research: The effects of diversity faultlines on conflict and performance," *Group Decis. Negot.*, vol. 12, no. 3, pp. 217–241, 2003.

[12] K. Bezrukova, K. A. Jehn, E. L. Zanutto, and S. M. B. Thatcher, "Do Workgroup Faultlines Help or Hurt? A Moderated Model of Faultlines, Team Identification, and Group Performance," *Organ. Sci.*, vol. 20, no. 1, pp. 35–50, Aug. 2008.

[13] B. Meyer and A. Glenz, "Team faultline measures: A computational comparison and a new approach to multiple subgroups," *Organ. Res. Methods*, vol. 16, no. 3, pp. 393–424, 2013.

[14] B. Meyer, A. Glenz, M. Antino, R. Rico, and V. González-Romá, "Faultlines and subgroups: A meta-review and measurement guide," *Small Group Res.*, vol. 45, no. 6, pp. 633–670, 2014.

[15] A. M. Carton and J. N. Cummings, "A theory of subgroups in work teams," *Acad. Manage. Rev.*, vol. 37, no. 3, pp. 441–470, 2012.

[16] C. N. Macrae, G. V. Bodenhausen, and A. B. Milne, "The dissection of selection in person perception: inhibitory processes in social stereotyping.," *J. Pers. Soc. Psychol.*, vol. 69, no. 3, p. 397, 1995.

[17] S. Roccas and M. B. Brewer, "Social identity complexity," *Personal. Soc. Psychol. Rev.*, vol. 6, no. 2, pp. 88–106, 2002.

[18] H. A. Schroth, G. Corniola, and M. Voit, "Teaching materials for negotiations and decision making." Northwestern University, Dispute Resolution Research Center.

[19] B. Meyer and A. Glenz, "Calculating diversity faultlines with the asw. cluster package in R: A step-by-step guide for beginners."

[20] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains," *J. Res. Personal.*, vol. 37, no. 6, pp. 504–528, 2003.

[21] S. Aslani *et al.*, "Dignity, face, and honor cultures: A study of negotiation strategy and outcomes in three cultures," *J. Organ. Behav.*, vol. 37, no. 8, pp. 1178–1201, 2016.

[22] M. J. Gelfand *et al.*, "Differences between tight and loose cultures: A 33-nation study," *Science*, vol. 332, no. 6033, pp. 1100–1104, 2011.

[23] L. R. Goldberg *et al.*, "The international personality item pool and the future of public-domain personality measures," *J. Res. Personal.*, vol. 40, no. 1, pp. 84–96, 2006.