

---

# Social Category Salience Moderates the Effect of Diversity Faultlines on Information Elaboration

Small Group Research  
42(3) 257–282  
© The Author(s) 2011  
Reprints and permission:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1046496411398396  
<http://sgr.sagepub.com>



Bertolt Meyer<sup>1</sup>, Meir Shemla<sup>2</sup>,  
and Carsten C. Schermuly<sup>3</sup>

## Abstract

Faultlines—hypothetical dividing lines splitting a group into homogeneous subgroups based on the distribution of demographic attributes—are frequently assumed to be detrimental to group outcomes because they operationalize social categorizations. However, a literature review indicates that this is not always the case. We argue that diversity faultlines and social categorizations are not necessarily the same and that the effect of diversity faultlines is moderated by perceived social categorizations. To test this proposition, we assigned 172 participants to groups of four. Participant gender, bogus personality feedback, seating position, and colored cards were employed to create two diversity faultline conditions (weak and strong faultline). Groups worked on the Survive in the Desert task, and their interactions were coded with the discussion coding system (DCS). Social categorizations were elicited using a newly developed measure that requires participants to specify subjectively perceived salient categories. Participants stated many social

---

<sup>1</sup>Universität Zürich, Zurich, Switzerland

<sup>2</sup>Technische Universität Dresden, Dresden, Germany

<sup>3</sup>Technische Universität Braunschweig, Braunschweig, Germany

Editor's Note: This article is part of the SGR special section: Research Presented at the 2010 Conference of the Interdisciplinary Network for Group Research.

## Corresponding Author:

Bertolt Meyer, Psychologisches Institut der Universität Zürich, Binzmühlestrasse 14/15,  
CH-8050 Zurich, Switzerland  
Email: [bmeyer@sozpsy.uzh.ch](mailto:bmeyer@sozpsy.uzh.ch)

categories that were unrelated to surface-level characteristics frequently employed in diversity research. In line with our hypotheses, social category salience moderated the effect of faultline strength on elaboration. Elaboration was most intense in strong faultline groups that had low levels of category salience. Elaboration was positively related to performance. Theoretical and practical implications are discussed.

## **Keywords**

Subgroups, Diversity, Faultlines, Social Categorization

As teamwork has become one of the most common forms of organizational collaboration (Salas, Cooke, & Rosen, 2008), we need to understand how group diversity relates to group processes and performance. For a long time, a dualistic view dominated diversity research (for reviews, see Joshi & Roh, 2008; van Knippenberg & Schippers, 2007; Williams & O'Reilly, 1998). Studies employing deep-level diversity ("differences among [group] members' attitudes, beliefs, and values" [Harrison, Price, & Bell, 1998, p. 97]) and surface-level diversity ("overt, biological characteristics that are typically reflected in physical features" [Harrison et al., 1998, p. 97]) usually find a negative relationship between surface-level and deep-level diversity on one side and group task performance on the other (e.g., Harrison et al., 1998; Harrison, Price, Gavin, & Florey, 2002; Zellmer-Bruhn, Maloney, Bhappu, & Salvador, 2008). These studies follow the social categorization perspective in diversity research (Williams & O'Reilly, 1998) in which a categorization of other people as either similar to the self (the ingroup) or different from the self (the outgroup) on salient (i.e., perceived) social categories entails intergroup bias—a more favorable evaluation of ingroup members than outgroup members. Intergroup bias is assumed to result in negative consequences for group social integration (i.e., decreased identification and commitment: Harrison et al., 1998, 2002; van Knippenberg, De Dreu, & Homan, 2004) that ultimately decrease task performance.

Another strand of research argues that informational diversity—differences among group members' task-relevant knowledge and perspectives—is beneficial for group performance (Horwitz & Horwitz, 2007; van Knippenberg et al., 2004). A heterogeneous knowledge base increases the likelihood that one group member finds a solution to a problem (e.g., Gruenfeld, Mannix, Williams, & Neale, 1996; Gruenfeld, Martorana, & Fan, 2000; Meyer & Scholl, 2009; Wittenbaum & Stasser, 1996). Differences in perspectives will

also prevent the group from reaching a quick consensus and will increase group-level information processing as group members try to understand the divergent positions (Meyer & Scholl, 2009; van Knippenberg et al., 2004). Summarizing these assumptions, demographic and deep-level diversity are thought to be inherently unfavorable (but see McLeod, Lobel, & Cox, 1996 for an exception), whereas informational and task-related diversity are thought to be inherently favorable.

The theoretical distinction between inherently positive and negative types of diversity was not supported empirically (van Knippenberg et al., 2004). The accumulation of mixed findings and null relationships between diversity and team outcomes (e.g., Horwitz & Horwitz, 2007) led to the development of the categorization-elaboration model (CEM) of group performance (van Knippenberg et al., 2004; van Knippenberg & Schippers, 2007). The CEM combines the predictions of the social categorization perspective with the predictions of the information/decision-making approach. According to the CEM, any diversity dimension can elicit both information/decision making and social categorization processes (van Knippenberg et al., 2004). The CEM assumes that diversity does not automatically lead to intergroup bias. Only if diversity leads to social categorizations on salient social categories, and these lead to intergroup bias, do diversity's negative affective consequences diminish the potentially positive relationship between diversity and outcomes, for example, decision quality.

Thus, diversity will have negative effects on performance, if perceived social categorizations lead to intergroup bias and its associated unfavorable affective reactions, such as increased conflict and decreased team social integration. These unfavorable affective reactions will reduce a potentially positive relationship between diversity and the elaboration of task-relevant information. If diversity does not lead to social categorizations to begin with, it can increase group performance by increasing the elaboration of task-relevant information and perspectives. Therefore, social categorizations on salient social categories and their impact on team social integration (e.g., cohesion, conflict, and identification) moderate the relationship between actual diversity and task performance. This relationship is mediated by the elaboration of task-relevant knowledge. Elaboration, defined as "the exchange of information and perspectives, individual-level processing of the information and perspectives, the process of feeding back the results of this individual-level processing into the group, and discussion and integration of its implications" (van Knippenberg et al., 2004, p. 1011), is seen as one of the core processes driving group decision quality and creativity in nonroutine tasks (van Knippenberg et al., 2004; Homan, van Knippenberg, van Kleef, &

De Dreu, 2007a, 2007b; Meyer & Schermuly, IN PRESS). As this definition of elaboration somewhat conflates cognitive and communication processes, and as other studies (e.g., Homan et al., 2007a) operationalized this concept with observational measures, we regard information elaboration as the act of exchanging, discussing, and integrating information and perspectives through verbal communication.

A divide between subgroups that is based on the group's diversity can also be referred to as a faultline, the dividing line that splits a group into relatively homogeneous subgroups based on the group members' alignment along multiple attributes (Bezrukova, Jehn, Zanutto, & Thatcher, 2009). The aim of this study is twofold. First, following the assumptions underlying the CEM model, we argue that faultlines can have positive or negative effects depending on a specific moderator, namely, social categorization. We intend to show that demographic faultlines can have a positive effect on an outcome variable, if social categorizations are low. Second, by illustrating the advantages of eliciting social categorizations using a measure that does not restrict participants' responses to certain types of social categories chosen by the researcher, we intend to scrutinize central theoretical questions about measuring group diversity and the relationship between objective and subjective diversity.

## **Diversity Faultlines**

One precondition for the salience of social categories is the extent to which observed similarities and differences between people or their actions are perceived as correlated with a division into social categories (Oakes, Turner, & Haslam, 1991; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). A faultline is a hypothetical dividing line that splits a group into relatively homogeneous subgroups based on the group members' alignment along multiple attributes (Bezrukova et al., 2009). Since the development of the faultline construct by Lau and Murnighan (1998) several studies have examined its impact on outcome variables (e.g., Meyer & Schermuly, IN PRESS; Bezrukova et al., 2009; Homan et al., 2007a, 2007b; Lau & Murnighan, 2005; Phillips, Mannix, Neale, & Gruenfeld, 2004; Rico, Molleman, Sánchez-Manzarnares, & van der Veegt, 2007; Thatcher, Jehn, & Zanutto, 2003). The results of these studies are mixed, with Thatcher et al. (2003) finding diversity faultline strength related curvilinearly to relationship conflict, process conflict, group morale, and group performance, Phillips et al. (2004; Experiment 2) uncovering no significant difference in the performance between groups with a strong faultline and groups with weak faultlines, and Lau and Murnighan (2005) noting that strong faultlines

resulted in more intragroup conflict and lower satisfaction than weak faultlines. Furthermore, in several recent studies, the relationship between faultline strength and team outcomes was determined by moderators such as the convergence between faultlines and information distribution (Homan et al., 2007a, 2007b) and the basis through which faultlines were defined, that is, social category versus informational. Finally, Meyer and Schermuly (IN PRESS) showed that demographic faultlines can have a positive impact on team performance, if teams are motivated and their diversity beliefs fit their actual level of diversity. Similarly, in a theoretical paper, Nishii and Goncalo (2008) argue that demographic faultlines have a positive influence on team creativity because faultlines can alert group members to the differences within the group and this makes them more open to new ideas from other group members.

One reason for the inconsistent results summarized above may lie in the fact that the “hypothetical dividing line” (Bezrukova et al., 2009, p. 35) constituting a diversity faultline does not always lead to an actual (i.e., perceived) and/or meaningful dividing line in the group. Those studies that did find a negative effect of demographic diversity faultlines on team outcomes (e.g., Homan et al., 2007a, 2007b) employed a manipulation check showing that participants in the strong faultline condition perceived their group as more diverse. Studies that did not find a negative effect of faultlines did not employ such a check (Phillips et al., 2004; Thatcher et al., 2003). Thus, in these studies at least, the hypothetical divide in the groups may not have translated into an actual one and strong faultlines did not lead to salient (i.e., perceived) categorizations on the categories employed for faultline creation or measurement. This argument is in line with findings by Oetzel (2001), who showed that perceptions of group members can have a stronger influence on group outcomes than compositional measures.

Furthermore, conceptual considerations support the idea that faultlines are not always salient. As outlined above, diversity faultlines operationalize potential comparative fit—the extent to which observed similarities and differences between people or their actions are perceived as correlated with a division into social categories (Turner et al., 1987). However, social category salience is determined by an interaction of comparative fit, normative fit, and accessibility (Oakes et al., 1991; Turner et al., 1987; van Knippenberg et al., 2004). *Normative fit* refers to the extent to which a categorization is meaningful to the group member (van Knippenberg et al., 2004), whereas *accessibility* refers to the readiness of a category to become activated (Oakes et al., 1991). Accessibility also reflects the responsiveness of the categorization process to the context of the perceiver (Oakes et al., 1991). Thus, if the

categories used to create a faultline are not meaningful to the group member in the task situation, the group member will possibly notice differences but will not categorize group members according to these differences (van Knippenberg et al., 2004). Therefore, perceived social categorizations should be elicited in conjunction with diversity faultlines.

The CEM assumes that comparative fit is one of the drivers of both the potentially positive and negative effects (van Knippenberg et al., 2004, p. 1011). Thus, faultlines are neither inherently good nor bad, and a distinction between desirable informational faultlines and undesirable demographic faultlines (see above) is not warranted. Instead, in line with the arguments of Phillips et al. (2004) and Nishii and Goncalo (2008), faultlines can alert group members to differences within the team without necessarily leading to social categorizations on situationally relevant (i.e., accessible) categories. If social categorizations on relevant and accessible categories takes place, these can lead to intergroup bias, which moderates the relationship between comparative fit and the elaboration of task-relevant information (van Knippenberg et al., 2004). Thus, salient social categorizations form the basis for a potential moderation of the association between comparative fit and the elaboration of task-relevant information. We thus postulate the following:

*Hypothesis 1:* The relationship between faultline strength and elaboration of task-relevant information is moderated by the intensity of salient social categorizations.

This hypothesis can explain the inconsistent findings of the diversity faultline literature reported above. It also extends the notion in the CEM that any form of diversity can have positive as well as negative effects to the faultline concept. Before presenting an experiment testing this hypothesis, we outline the issues that are associated with the measurement of social categorization and develop a novel measure for its elicitation.

## Measuring Social Categorizations

As our hypothesis emphasizes the centrality of social categorization in understanding the impact of team diversity, we find it critical to elaborate on the conceptualization and measurement of this construct.

A review of the literature teaches us that several studies investigated the effects of perceived diversity (i.e., social categorization) on group outcomes (Harrison et al., 2002; van Dick, van Knippenberg, Hagele, Guillaume, & Brodbeck, 2008; Zellmer-Bruhn et al., 2008). These studies used fixed sets of categories (age, ethnicity, personality, communication style, work ethics,

attitudes about school and education, and project goals, among others), chosen by the researchers, for eliciting measures of perceived intragroup differences. The researchers thus implicitly assumed that these are the salient categories on which their participants make social comparisons. The reasons stated for selecting these categories were either left unstated (van Dick et al., 2008), included because they “have a fairly long history of study” (Harrison et al., 2002, p. 1031), mentioned in pilot interviews with students from the sampled population (Harrison et al., 2002), or “salient to the MBA student population” (Zellmer-Bruhn et al., 2008, p. 42), although no explanation is given how salience was determined or how they were “relevant to the tasks required of these MBA ‘core’ teams” (Zellmer-Bruhn et al., 2008, p. 42). The heterogeneity of the constructs surveyed along with their scarce reasoning can be interpreted as a sign of arbitrariness in the operationalization of diversity perceptions. The number of possible attributes appears infinite.

We argue that simply selecting certain diversity attributes for measuring social categorizations is too narrow and that a measure of perceived diversity must be flexible enough to incorporate the specific social categories that group members employ to form an impression. “A specific situation and social construction may make salient certain attributes *whether or not they are relevant to the task* [italics added]. If salient, these distinctions, regardless of how task-relevant they are, may lead to in-group/out-group distinctions and potentially affect group functioning” (Williams & O’Reilly, 1998, p. 81). Thus, category salience is highly situational and is not necessarily determined by a category’s task relevance. It is unknown whether the specific attributes stated above are the attributes that become salient in a task situation. It is also unclear whether attributes that researchers choose to elicit are the attributes that affect group functioning (the most). We are aware of one other study that follows these notions and that investigated perceived social categories without restraining participants to a set of predetermined categories. Oosterhof, van der Vegt, van de Vliert, Sanders, and Kiers (2009) asked 80 members of different teams of a nonprofit organization about attributes that they perceive as differentiating them from their colleagues. In response, participants named 497 dissimilarity attributes—an average of almost three attributes per colleague. In the present study, we adopt this approach and extend it by asking participants about categories that differentiate between members of their work team and by adding a numerical scale for intensity perceptions to this qualitative measure (see below). To assess the validity of such a measure, we postulate the following:

*Hypothesis 2:* Social category salience intensity perceptions for a specific stated social category are related to actual within-group differences with regard to that category.

As comparative fit is one determinant of social category salience (see above), faultlines should also influence salient social categorizations. We thus postulate the following:

*Hypothesis 3:* Salient social categorizations are positively related to the strength of a diversity faultline within the group.

In the following section, we present an experiment that tests the hypotheses outlined using a newly constructed measure for social category salience.

## Method

### Sample

The participants, who received a payment of CHF40 (approximately EUR30) for participating, were 172 students (86 women) from two different major Swiss universities in Zurich. On average, participants were 22.2 years old ( $SD = 2.70$ ). Sixty-four participants were majoring in mathematics or natural sciences, 63 in social sciences, 25 in engineering, 6 in business, 4 in law, 4 in architecture, and 3 in medicine. Three participants did not state their field of study. All participants were of White ethnicity. In terms of nationality, 141 were Swiss, 25 German, and 2 were Austrian. There was also one student from Italy, one from the Netherlands, the United States, and one from Turkey. All participants were fluent in German. The study was advertised as a study on group dynamics. Participants were randomly assigned to 43 groups, each consisting of two male and two female participants.

### Procedure

Before the experiment took place, participants completed an online questionnaire on their demographics that also included a personality measure (see below). Participants were then invited to participate in an experimental session, where they were greeted by a female experimenter and seated in a room with audio and video recording facilities. Participants were asked to work on a modified desert survival task that we obtained from Homan et al. (2007a) and translated into German. In the task, participants first read some general information on strategies for surviving in the desert individually. They were then asked to brainstorm on items that are useful for surviving in the desert, that can be carried, and that meet certain specifications and to then make a list of useful items. This group discussion was recorded on video. Each item



had to be justified in writing. At the end of the exercise, which had a 30-min time limit, participants were asked to write down the three most important items that they would like to bring along to the desert. Participants were informed that the members of the three best groups would earn another cash prize ranging from 50 to 200 CHF per person.

We chose this particular task for several reasons. First, the CEM's predictions with regard to the potential positive effects of diversity apply to creative and/or nonroutine tasks. The process of brainstorming for items introduces an aspect of creativity to the task as brainstorming is considered a creative task (Diehl & Stroebe, 1987). Second, the CEM's predictions only apply to tasks in which the elaboration of information can potentially impact task performance (van Knippenberg et al., 2004). Third, this particular version of the desert survival task was employed in two other studies that examined the impact of faultlines on team performance and information elaboration (Homan et al., 2007a, 2007b). Hence, the employment of this task ensures comparability of our results with other studies on team diversity in the context of the CEM.

After reading the instructions individually but before they started to work on the task, the participants completed a questionnaire on task motivation. After participants had completed the task, they worked on a questionnaire on perceived social category salience. Once that was completed, participants worked on scales from the Berlin Intelligence Structure (BIS) Test (Jäger, Süß, & Beauducel, 1997) that were employed for assessing the validity of the social category salience measure (see below). Afterward, participants were thanked and debriefed. The entire experiment lasted approximately 1.5 h.

*Faultline manipulation.* In the experimental sessions, the strength of the groups' faultline was varied experimentally through participants' gender, through their seating position, and through bogus personality feedback. The participants were seated at a large table with two participants facing the two other participants. In the strong faultline condition, the two female participants sat on one side of the table and the two male participants sat on the other side of the table. In the weak faultline condition, a male and a female participant sat on both sides of the table. After participants were seated, they received bogus feedback on their dominant personality type, which was deemed credible because of the online personality pretest. In the strong faultline condition, both women were told that they were V-type personalities and the two men were told that they were O-type personalities. In the weak faultline condition, one woman on one side of the table was told that she was O and the woman on the other side was told that she was V; the men also received unequal feedbacks. Colored cardboard signs in front of the participants,

bearing their number for identification purposes, increased the salience of the faultline in the strong faultline condition: The two women had red signs, the men had blue ones. In the weak faultline condition, there was a blue and a red sign on either side of the table. The success of the manipulation was assessed by two items on the posttask questionnaire (see below).

## Measures

*Demographic diversity.* Age diversity of groups was determined with the group-level standard deviation. Diversity with regard to participants' nationality was determined by computing a Blau index (Blau, 1977) for nationality.

*Personality.* The Big Five personality traits were measured using the German NEO-FFI questionnaire (Körner et al., 2008). It contained the five scales with six items each: Neuroticism ( $\alpha = .82$ ), Extraversion ( $\alpha = .74$ ), Openness ( $\alpha = .74$ ), Agreeableness ( $\alpha = .64$ ), and Conscientiousness ( $\alpha = .74$ ).

*Information elaboration.* The verbal interactions among the group members were coded using the discussion coding system (DCS; Schermuly & Scholl, IN PRESS). The DCS is an instrument for analyzing larger samples of group interactions. It was designed in a way that makes it applicable to a wide range of situations, so that findings from different studies become comparable and standards or norms can be developed. For each speech act, the DCS captures its accompanying interpersonal affect, its function, and its responses. It is adapted to the sequential, vertical, and reciprocal nature of interaction (Boos, 1995) and is thus suitable for capturing group processes. The DCS unitizes the group interaction into individual speech acts. Speech acts can be classified as belonging to certain categories (Bales, 1950; Fisch, 1994). In the case of the DCS, a *speech act* refers to a sentence-like unit that can be attributed to one of the three main categories: social-emotional statement (differentiated as positive or negative), statement with regard to the content of the task, or regulatory statement. For each of these three main categories, the two minor categories proposal and question can be coded.

The unitization of the speech acts is based on a set of hierarchical rules (Schermuly & Scholl, IN PRESS). Coders code a new act, if (a) the speaker changes, or (b) the speaker starts to address a different person, or (c) the speaker changes from one main category into another (e.g., speaker starts with a regulation and then switches to a socioemotional statement), or (d) the speaker states a new minor category, or (e) the speaker speaks for more than 30 s, or (f) the speaker stays in the same functional domain but the main argument explicitly changes. Two independent coders were trained on these rules. Afterward, they coded three randomly chosen discussions each. Across these,

1,015 speech acts were identified by at least one of the two coders; 95.7% of these speech acts were identified by both coders in the same manner. Hence, we judged the unitization to be reliable. The remaining discussions were subsequently sequenced into speech acts by these two coders. The sequenced discussions were then supplied to a second coding procedure, during which the functions of the speech acts were categorized into major and minor categories as described above by a different group of six student coders who were blind to experimental conditions and who had considerable experience on the usage of the DCS from another study. In this way, the same speech act could receive two different ratings from two different coders in the second procedure, but coders could not disagree on the unitization of speech acts.

To obtain a measure for the depth of information elaboration, we followed the principles outlined in Homan et al. (2007a): Each speech act that was coded as a statement with regard to content received an elaboration score of 1. If other group members “reacted to it (e.g., by saying something like ‘OK’ or by nodding) but after this the group still failed to ask questions about it or integrate it with the other information” (Homan et al., 2007a, p. 1193), then the act received a score of 2. A score of 3 was given, if a statement was a reaction to a previous statement indicating that it had been processed by asking a question about it, by drawing a conclusion from it, or by combining it with another piece of information. This elaboration score was divided by the total number of speech acts in a group, resulting in a maximum elaboration ratio of 3 per group member. The mean elaboration ratio per group member was 0.30 ( $SD = .20$ ).

One video could not be coded because of technical issues. The remaining 42 videos were assigned randomly to coders, and 20 videos were coded twice by two different coders. The resulting intraclass correlation (ICC) for the weighted elaboration measure was .72. To shed some light on the validity of the elaboration measure, we correlated the group-mean elaboration of task-relevant information with group performance. In line with the CEM’s predictions, we found a positive relationship,  $r = .32, p = .04$ .

**Task performance.** Quantitative and qualitative performance was determined using the coding scheme developed by Homan et al. (2007a): Quantitative performance was determined by calculating the mean score per item for each group. In the original task, 12 categories of items were distinguished. On the basis of these categories, we developed a coding scheme by which performance could be calculated. Better items (i.e., items that were ranked higher in the expert ranking reported by Johnson and Johnson [1982]) received a higher score, with the highest possible score being 12 points. For example, when the group decided that it would bring a magnetic compass to the desert, it received a score of 1 because the instructions clearly indicated that groups

should not walk (Homan et al., 2007a, p. 1192). Quality of performance was determined by dividing the quantitative score by the number of items generated. This ensures that groups that devised a lot of low-quality items would not get higher scores than groups that came up with fewer high-quality items (Homan et al., 2007a). Two independent raters coded the items generated by the groups. Interrater reliability was computed with the ICC, which turned out to be good (ICC = .71).

*Social category salience.* The participants were asked to write down three features that in their view differentiated group members from each other. In Oosterhof et al. (2009), participants also reported an average of three differences per relationship. For each feature, participants were asked how strong they perceived this difference in the group to be (from 1 = *not strong* to 5 = *very strong*). These three ratings do not operationalize a single underlying construct as they refer to different attributes. However, they are causal indicators (MacCallum & Browne, 1993) of social category salience as they give rise to an overall strength of perceived salience: A participant who perceives three attributes to a very strong extent perceives stronger differences than a participant who perceives three attributes with a weak strength. We therefore averaged these three items into one measure.

To assess the validity of this measure, we assessed further deep-level diversity attributes (Harrison et al., 1998) that van Knippenberg et al. (2004) mentioned in the context of the CEM: task motivation, cognitive ability, and creativity. Task motivation was measured with the items from two scales from the Questionnaire on Current Motivation (Rheinberg, Vollmeyer, & Burns, 2001): Interest (“I like problems of this kind”; “After reading the instructions, the task seems to be very interesting”; “For tasks like this I don’t need a reward, they are lots of fun anyhow”; “I would work on this task even in my free time”;  $\alpha = .82$ ) and Challenge (“This task is a real challenge for me”; “I am eager to see how I will perform in this task”; “I am really going to try as hard as I can on this task”; “If I can do this task, I will feel proud of myself”;  $\alpha = .68$ ). To test whether these scales operationalize a higher order construct task motivation, we submitted the items to a confirmatory factor analysis. Each item was specified as an indicator for its corresponding scale. The two scales were again specified as indicators for a higher level factor representing task motivation. We employed the MPlus software (Muthén & Muthén, 2008) with the MLMV (Mean- and Variance-Adjusted Maximum Likelihood) estimator that takes the small sample size into account. According to the criteria specified by Schermelleh-Engel, Moosbrugger, and Müller (2003), the model showed an acceptable model fit,  $\chi^2 = 28.33$ ,  $df = 15$ ,  $p = .02$ , Comparative Fit Index = 0.954, RMSEA = .072. The higher order factor task motivation showed standardized loadings of .85 on Interest and of

.89 on Challenge. Therefore, we combined the items into one measure for task motivation ( $\alpha = .83$ ) and employed the within-group standard deviation as a measure for motivational diversity.

For capturing the diversity with regard to participants' cognitive ability, they completed two scales from the short version of the BIS Test (Jäger et al., 1997). The first one, a reasoning scale, consisted of six speed tests: number sequence completion, pattern analogies, word analogies, fact-opinion distinction, mathematical estimation, and figure completion. The normed scores of these six tests are averaged for the total reasoning score ( $\alpha = .73$ ). Again, the within-group standard deviation was used for assessing group diversity with regard to this measure.

Finally, we assessed participants' creativity with a scale on figural idea flexibility from the BIS Test (Jäger et al., 1997). The test consists of a series of circles that participants have to complete into drawings. Two independent raters scored the drawings for the number of superior categories employed according to the BIS coding scheme. Interrater reliability was good, ICC = .86. Raw scores were transformed into norm scores following the BIS manual.

*Manipulation checks.* To assess the success of the faultline manipulation, each participant rated their perceived overall similarity to all other group members. If the faultline manipulation was successful, the group member on the same side of the table should receive a higher similarity rating under the high faultline condition than the two group members on the other side. We also included an item reading "Our group was strongly divided due to gender, personality, and seating" with a scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

## Results

### *Manipulation Checks*

We compared the general similarity ratings for group members on the same side of the table with the similarity ratings for group members on the other side of the table. We tested their differences with a mixed (between and within) ANOVA. It employed the similarity rating for the member on the same side of the table and the average similarity rating for the two members on the opposite side of the table as within-subject factor and the experimental condition (strong faultline / weak faultline) as between-subject condition. It revealed the expected within-and-between interaction of Table side  $\times$  Experimental condition,  $F(1,338) = 23.49, p < .001, \eta = .25$ . Neither the main effect of table side,  $F = 1.77, p = .184$ , nor the main effect of faultline condition ( $F < 1$ ) reached levels of statistical significance.

Under the strong faultline condition, the mean similarity rating for the member on the same side of the table was 3.39 ( $SD = 0.98$ ), whereas the mean average similarity rating for the two other group members was 2.92 ( $SD = .62$ ,  $d = .43$ ). Under the weak faultline condition, participants reported a lower similarity for the member on the same side of the table ( $M = 2.95$ ,  $SD = 0.86$ ) than for the two members on the other side of the table ( $M = 3.31$ ,  $SD = .62$ ,  $d = -.34$ ).

Under the strong faultline condition, mean agreement with the item "Our group was strongly divided due to gender, personality, and seating" was significantly higher,  $M = 2.27$ ,  $SD = 1.19$ , than under the weak faultline condition,  $M = 1.86$ ,  $SD = 0.89$ ,  $t(170) = 2.58$ ,  $p = .011$ ,  $d = .39$ . The manipulation was deemed successful.

### Level Issues

The data obtained in this experiment are of a multilevel structure: Personality, motivation, information elaboration, social category salience, reasoning ability, and creativity were elicited at the individual level, whereas the faultline manipulation took place at the group level. Surface-level and deep-level diversity were also computed at the group level. Consequently, we employed multilevel analysis to account for group effects on the dependent and moderating variables (Gelman & Hill, 2006). All multilevel analyses were computed with the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core team, 2009) of the R software (R Development Core Team, 2008). Following recommendations by Bliese (2009), we tested the appropriateness of a multilevel model for the employed dependent variables in their respective analyses.

### Faultlines, Social Category Salience, and Elaboration

To test the assumption that the intensity of perceived social categorizations moderates the relationship between faultline strength and the elaboration of task-relevant information (Hypothesis 1), we conducted a moderation analysis. It regressed elaboration on social category salience, on faultline strength, and on their interaction. We also controlled for the measures of surface- and deep-level diversity. An ICC(1) of 0.57,  $F(41, 126) = 6.27$ ,  $p < .001$ , indicated that the elaboration of task-relevant information was of a nested nature, warranting a multilevel regression. An ICC(2) of 0.85 showed that the group mean of the elaboration measure was reliable. Thus, multilevel modeling was warranted (Bliese, 2009). Table 1 presents the multilevel regression model.

**Table 1.** Multilevel Regression of Information Elaboration on Social Category Salience, Faultline Strength, Their Interaction, Surface-Level Diversity, and Deep-Level Diversity

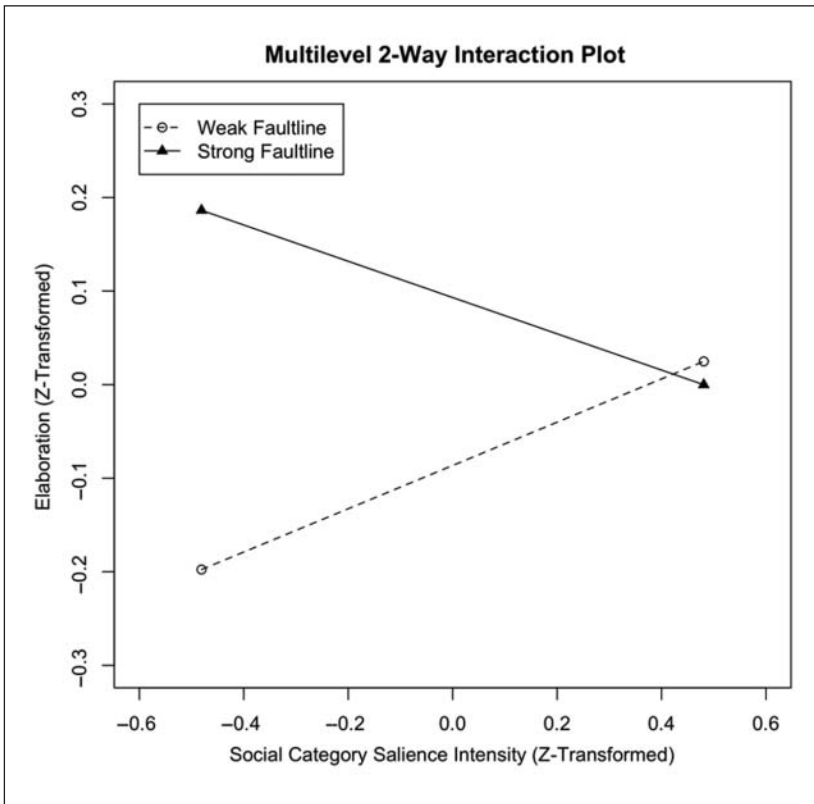
Variable	Parameter estimate	SE	df	t
(Intercept)	0.00	0.10	124	0.03
CS	0.02	0.11	124	0.17
F	0.47	0.62	30	0.75
Age (SD)	0.00	0.13	30	-0.02
Nationality (Blau index)	-0.27	0.12	30	-2.31*
Extraversion (SD)	0.29	0.11	30	2.58*
Agreeableness (SD)	-0.25	0.12	30	-2.07*
Neuroticism (SD)	-0.04	0.12	30	-0.35
Openness (SD)	-0.12	0.12	30	-0.96
Conscientiousness (SD)	0.24	0.12	30	2.03 <sup>+</sup>
Motivation (SD)	0.18	0.12	30	1.48
Reasoning (SD)	-0.02	0.11	30	-0.16
Creativity (SD)	0.07	0.12	30	0.54
CS × F	-1.11	0.56	124	-1.98*

Note: CS = social category salience; F = faultline strength.

<sup>+</sup> $p < .10$ , \* $p < .05$  (two-tailed)

A comparison of the  $-2$  log likelihood ( $-2LL$ ) of the model with a null model without predictors revealed a marginally significant decrease,  $\Delta -2LL = 23.54$ ,  $df = 15$ ,  $p = .073$ . Due to the conservativeness of the test (Thomas, Bliese, & Jex, 2005), this decrease illustrates the informational value of the model. We therefore determined between-group and within-group variance explanations by comparing the intercept and residual variance of the null model with the calculated model (Bliese, 2009). The model accounted for 43.4% of intergroup variance and for 3.3% of the observed within-group differences.

In support of the hypothesis, the model revealed a significant cross-level interaction. To facilitate interpretation, we plotted a multilevel interaction plot with the tools provided by Preacher, Curran, and Bauer (2006); see Figure 1. The plot revealed that groups with strong faultlines and low values of perceived social category salience intensity exhibited the highest levels of elaboration. If group members of groups with strong faultlines reported high levels of category salience, information elaboration decreased. This finding supports a central notion of the CEM: If comparative fit is associated with social categorizations, elaboration decreases. If it is not, elaboration increases.



**Figure 1.** Cross-level moderation effect of social category salience intensity on the relationship between diversity faultlines and the elaboration of task-relevant information

The plot revealed another effect for groups in the weak faultline condition: If faultline strength and social category perceptions were low, group members elaborated the least. However, if the weak faultline manipulation coincided with perceptions of social category salience, the elaboration of task-relevant information increased.

An independent samples *t* test revealed that group members in the weak faultline condition who reported below-average social category salience exhibited less elaboration ( $M = -0.21$ ) than group members in the strong faultline condition who reported below-average social category salience,  $M = 0.21$ ,  $t(76) = -1.90$ ,  $p = .03$  (one-tailed). For group members who



reported a social category salience intensity above the mean, the difference was not significant,  $t < 1$ .

### *Social Category Salience Construct Validity*

We evaluated the 515 diversity attributes that were stated by the participants in response to the social category salience items. Two independent coders classified the attributes as either deep level or surface level (Harrison et al., 2002). Only 27 attributes were classified as surface-level diversity attributes. These were 17 mentions of gender, 5 mentions of age differences, 4 mentions of clothing style or physical appearance, and 1 mention of the bogus personality type. The remaining attributes spanned diverse deep-level attributes, such as creativity (the highest mentioned with 64 mentions), humor, teamwork, motivation, personality features, discussion style, goal orientation, imaginativeness, fantasy, directness, activity, friendliness, talkativeness, and many more.

We calculated the group mean score for the social category salience based on the mentions of creativity (available for 36 groups). It correlated with the group-level standard deviation of the employed objective creativity measure,  $r = .33, p < .05$ . Thus, those participants who perceived creativity as a salient social category tended to perceive the actual group-level diversity of creativity, which indicates construct validity of the measure. Hypothesis 2 was thus supported by the data.

### *Determinants of Social Category Salience*

To test whether social category salience was of a hierarchical nature, we examined its interrater agreement with the  $r_{WG(j)}$  index (Cohen, Doveh, & Nahum-Shani, 2009) and group-level variance with the ICC(1) (Bliese, 2009; Shrout & Fleiss, 1979). The  $r_{WG(j)}$  of .81 indicated agreement (LeBreton & Senter, 2008), and an ICC(1) of 0.11,  $F(42, 129) = 1.51, p < .05$ , indicated that a significant amount of variance was determined by group membership.

To test whether faultlines, demographic diversity (i.e., surface-level diversity), and diversity in terms of personality, intelligence, and creativity (i.e., deep-level diversity) influenced salience of social categorizations as predicted by Hypothesis 3, we conducted a multilevel analysis (Gelman & Hill, 2006; see Table 2). All variables were Z transformed. To determine the explanatory power of the model, we compared its  $-2LL$  value to a null model that only contained a group-level intercept as a predictor (Bliese, 2009; Gelman & Hill, 2006). The decrease in the  $-2LL$  value between the

**Table 2.** Multilevel Regression of Social Category Salience on Faultline Strength, Surface-Level Diversity, and Deep-Level Diversity

Variable	Parameter estimate	SE	df	t
Faultline strength	-0.12	0.08	31	-1.48
Age (SD)	0.26	0.09	31	2.79*
Nationality (Blau index)	0.11	0.08	31	1.31
Extraversion (SD)	0.15	0.08	31	1.82 <sup>+</sup>
Agreeableness (SD)	-0.13	0.09	31	-1.50
Neuroticism (SD)	-0.01	0.09	31	-0.06
Openness (SD)	0.07	0.09	31	0.78
Conscientiousness (SD)	-0.11	0.09	31	-1.23
Motivation (SD)	-0.06	0.09	31	-0.68
Reasoning (SD)	0.04	0.08	31	0.56
Creativity (SD)	0.07	0.09	31	0.80

+ $p < .10$ , \* $p < .05$  (two-tailed)

null model and the tested model was small and not significant,  $\Delta -2LL = 15.11$ ,  $df = 11$ ,  $p = .178$ . Thus, the tested model did not have a significantly higher informational value than a model without predictors. In summary, despite a small influence from age diversity, the intensity of perceived social categorizations was unaffected by the faultline manipulation and by available measures of surface-level and deep-level diversity. Hypothesis 3 was thus refuted.

## Discussion

Based on the CEM (van Knippenberg et al., 2004), we tested the assumption that the salience of social categories in work groups moderates the relationship between comparative fit—operationalized through diversity faultlines—and the elaboration of task-relevant information. We also tested how diversity faultlines influence the salience of social categories, which we elicited with a novel measure that consisted of open questions and quantitative data.

The results revealed two effects. First, the faultline manipulation, although successful, had no effect on the social category salience. This is not surprising: As stated in the introduction, the salience of social categories depends on comparative fit, normative fit, and accessibility of the social category (Oakes et al., 1991; Turner et al., 1987; van Knippenberg et al., 2004). Diversity faultline strength operationalizes only one of these three determinants of category salience, comparative fit. Thus, if the task situation makes other categories than the ones used for the faultline manipulation accessible, the comparative

fit of nonaccessible categories does not influence their salience. In the current case, the inspection of the social categories that were stated by the participants revealed that these were remarkably different from the surface-level diversity attributes that were employed to create the faultline. As stated in the Results section, the majority of the reported categories were related to specificities of the task, for example, creativity and imaginativeness. Thus, the task made other categories salient than those employed in the faultline manipulation (gender, seating, personality feedback) and hence made other categories accessible in the task situation. This finding shows that surface-level diversity attributes that a researcher chooses to elicit may not be the ones that are relevant to the group members in a specific task situation.

This finding has implications for the way that perceived diversity and demography are studied. Predefined categories may not capture what study participants would describe, if they were not limited to those categories. Perceived diversity can be something that is contextually specific, determined by the specificities of the perceiver and the situation. Questions with regard to age, gender, and ethnicity may be insufficient for capturing these specificities. Given the fact that most attributes studied in diversity research were not central for the participants of this study, our results indicate that there is room for new research that focuses on the diversity that is of subjective relevance to study participants. Methods relying on open questions can be employed for capturing these perceptions without sacrificing quantitative rigor.

Second, in support of our hypothesis, social category salience moderated the effect of the faultline strength on the elaboration of task-relevant information: Groups with strong faultlines and low social category salience elaborated most, whereas an increase in social category salience decreased the elaboration of task-relevant information for these groups. The successful manipulation check showed that groups in the strong faultline condition did in fact perceive their diversity to a larger extent than groups in the weak faultline condition did. If these perceptions of diversity did not co-occur with social categorizations, elaboration improved. This finding supports one of the core assumptions of the CEM: Comparative fit can increase elaboration, if social categorizations do not take place. On the same side, groups with low levels of comparative fit (i.e., potentially visible diversity: faultlines) and low levels of social categorizations elaborated considerably less. This might be due to the fact that visible differences in terms of comparative fit might pose a *conditio sine qua non* for the potentially positive effects of diversity.

Groups with high levels of perceived social categorizations elaborated at medium levels and did not exhibit different levels of elaboration with different faultline strengths. This shows that perceived social categorization is not

automatically detrimental either. It seems that perceived differences on accessible categories can be an advantage in the absence of faultlines but are detrimental in their presence. Thus, social categorizations and faultlines should always be examined together, and one measure cannot substitute the other, as they can be independent of each other (see above).

### *Limitations*

This study is not without limitations. For example, we did not take into account participants' diversity beliefs, which "may be defined as beliefs about the value of diversity to work group functioning" (van Knippenberg, Haslam, & Platow, 2007, p. 209). Other studies showed that diversity beliefs can influence the effect of diversity faultlines on outcome variables (e.g., Homan et al., 2007a, 2007b; Meyer & Schermuly, IN PRESS). In this study, groups with strong diversity faultlines and high levels of performance might also have held pro-diversity beliefs. Future studies should thus examine the relationship between diversity faultlines, social categorizations, and diversity beliefs simultaneously.

Furthermore, the observed positive effects of diversity faultline strength might have been caused by the short-lived nature of the group interaction. Joshi and Roh (2008, 2009) showed that diversity is more likely to have positive effects in short-lived groups and is more likely to have negative effects in groups that work together for longer periods of time (but see Watson, Kumar, & Michaelsen, 1993 for opposite effects). Thus, the current findings need to be replicated in field contexts where teams work together for longer than half an hour before dissolving.

### *Implications and Outlook*

The findings have several implications for research and practice. The results show that faultlines can alert group members to their differences (e.g., Phillips & Loyd, 2006; Phillips, Northcraft, & Neale, 2006) and lead to an increase in elaboration of task-relevant information. Thus, our findings add to other recent research showing that faultlines are not inherently detrimental (Meyer & Schermuly, IN PRESS; Nishii & Goncalo, 2008). Practitioners can make use of this finding by trying to establish contexts in which diversity faultlines can do good. These would require the prevention of social categorizations when faultlines are present. One possibility might lie in the establishment of prodiversity beliefs among team members (e.g., Homan et al., 2007a, 2007b; Stegmann & van Dick, 2009). As other recent research highlighted the role of

leadership in diverse teams (Kearney & Gebert, 2009), leadership behavior aimed at decreasing the salience of social category salience if strong faultlines are present might be useful. Further studies could also address the conceptual conflation of communication and cognitive processes in the concept of information elaboration by looking at the cognitive and communication processes of elaboration separately.

Finally, Joshi and Roh (2009) pointed out that the organizational context of diverse teams can influence the relationship between diversity and outcomes. The same could be true for faultlines: For example, diversity faultlines in ethnically diverse organizations or areas may have different effects than in teams that are embedded into homogeneous environments. Further research should therefore investigate the role of context on the effects of faultlines.

In summary, our findings demonstrate that diversity faultlines are neither inherently good nor bad. Under the given circumstances, that is, low levels of social category salience, they can have a positive effect. Thus, group constellations that underline visible differences may have a positive influence on the elaboration of task-relevant information if social categorizations do not occur. This finding also supports the CEM model of diversity and group performance and thus contributes to the effort to combine previously conflicting views in diversity research. Finally, the new measure employed for social category salience that combines open questions and quantitative measures challenges some of the assumptions of diversity research: Perceptions of salient social categories differed markedly from the social categories that are most frequently employed in diversity research (e.g., age, gender, and ethnicity). This indicates that the group process could be a factor that socially constructs categorization. Further studies on the relationship between social category salience and diversity perceptions appear warranted.

### **Acknowledgments**

The authors thank Anita Huber for conducting the experiments and Astrid Homan and Daan van Knippenberg for their valuable comments on earlier versions of this manuscript.

### **Declaration of Conflicting Interests**

The author(s) declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

### **Funding**

The author(s) disclosed that they received the following support for their research and/or authorship of this article: This study was financially supported by the University of Zurich, Zurich, Switzerland.

## References

- Bales, R. F. (1950). A set of categories for the analysis of small group interaction. *American Sociological Review*, *15*, 257-263. doi:10.2307/2086790
- 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*, 35-50. doi:10.1287/orsc.1080.0379
- Blau, P. M. (1977). *Inequality and heterogeneity*. New York, NY: Free Press.
- Bliese, P. D. (2009). *Multilevel modeling in R (2.3)* [Online manual]. Retrieved from [http://cran.r-project.org/doc/contrib/Bliese\\_Multilevel.pdf](http://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf)
- Boos, M. (1995). Die sequentielle Strukturierung sozialer Interaktion [The sequential structure of social interaction]. In W. Langenthaler & G. Schiepek (Eds.), *Selbstorganisation und Dynamik in Gruppen* (pp. 209-221). Münster, Germany: Lit Verlag.
- Cohen, A., Dovich, E., & Nahum-Shani, I. (2009). Testing agreement for multi-item scales with the indices  $r_{WG(j)}$  and  $AD_{M(j)}$ . *Organizational Research Methods*, *12*, 148-164. doi:10.1177/1094428107300365
- Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*, 497-509. doi:10.1037/0022-3514.53.3.497
- Fisch, R. (1994). Eine Methode zur Analyse von Interaktionsprozessen beim Problemlösen in Gruppen [A method for analyzing interaction processes in group problem solving]. *Gruppendynamik*, *25*, 149-168.
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
- Gruenfeld, D., Mannix, E., Williams, K., & Neale, M. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, *67*, 1-15. doi:10.1006/obhd.1996.0061
- Gruenfeld, D., Martorana, P., & Fan, E. (2000). What do groups learn from their worldliest members? Direct and indirect influence in dynamic teams. *Organizational Behavior and Human Decision Processes*, *82*, 45-59. doi:10.1006/obhd.2000.2886
- Harrison, D., Price, K., & Bell, M. (1998). Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. *Academy of Management Journal*, *41*, 96-107. doi:10.2307/256901
- Harrison, D., Price, K., Gavin, J., & Florey, A. (2002). Time, teams, and task performance: Changing effects of surface- and deep-level diversity on group functioning. *Academy of Management Journal*, *45*, 1029-1045. doi:10.2307/3069328
- Homan, A. C., van Knippenberg, D., van Kleef, G. A., & De Dreu, C. K. W. (2007a). Bridging faultlines by valuing diversity: Diversity beliefs, information elaboration,

- and performance in diverse work groups. *Journal of Applied Psychology*, 92, 1189-1199. doi:10.1037/0021-9010.92.5.1189
- Homan, A. C., van Knippenberg, D., van Kleef, G. A., & De Dreu, C. K. W. (2007b). Interacting dimensions of diversity: Cross-categorization and the functioning of diverse work groups. *Group Dynamics: Theory, Research, and Practice*, 11, 79-94. doi:10.1037/1089-2699.11.2.79
- Horwitz, S., & Horwitz, I. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 33, 987-1015. doi:10.1177/0149206307308587
- Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test* [Berlin Intelligence Structure Test] (*BIS-Test*). Göttingen, Germany: Hogrefe.
- Johnson, D. W., & Johnson, F. P. (1982). *Joining together: Group theory and group skills* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Joshi, A., & Roh, H. (2008). A contextual reexamination of work team diversity research: Review and future agenda. In K. W. Phillips (Ed.), *Research on managing groups and teams* (Vol. 11, pp. 27-53). Bingley, UK: JAI Press. doi:10.1016/S1534-0856(08)11002-7
- Joshi, A., & Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal*, 52, 599-627.
- Kearney, E., & Gebert, D. (2009). Managing diversity and enhancing team outcomes: The promise of transformational leadership. *Journal of Applied Psychology*, 94, 77-89. doi:10.1037/a0013077
- Körner, A., Geyer, M., Roth, M., Drapeau, M., Schmutzer, G., Albani, C., . . . Brähler, E. (2008). Persönlichkeitsdiagnostik mit dem NEO-Fünf-Faktoren-Inventar: Die 30-Item Kurzversion (NEO-FFI-30) [Personality assessment with the NEO Five-Factor Inventory: The 30-Item Short Version (NEO-FFI-30)]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 58, 238-245. doi:10.1055/s-2007-986199
- Lau, D., & Murnighan, J. (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review*, 23, 325-340. doi:10.2307/259377
- Lau, D., & Murnighan, J. (2005). Interactions within groups and subgroups: The effects of demographic faultlines. *Academy of Management Journal*, 48, 645-659.
- LeBreton, J., & Senter, J. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11, 815-852. doi:10.1177/1094428106296642
- MacCallum, R., & Browne, M. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, 114, 533-541.
- McLeod, P. L., Lobel, S. A., & Cox, T. H. (1996). Ethnic diversity and creativity in small groups. *Small Group Research*, 27, 248-264. doi:10.1177/1046496496272003

- Meyer, B., & Schermuly, C. C. (IN PRESS). When beliefs are not enough: Examining the interaction of diversity faultlines, task motivation, and diversity beliefs on team performance. *European Journal of Work and Organizational Psychology*.
- Meyer, B., & Scholl, W. (2009). Complex problem solving after unstructured discussion: Effects of information distribution and experience. *Group Processes & Intergroup Relations*, 12, 495-515. doi:10.1177/1368430209105045
- Muthén, L., & Muthén, B. (2008). MPlus version 5.1 [Computer Software]. Los Angeles, CA: Author.
- Nishii, L. H., & Goncalo, J. A. (2008). Demographic faultlines and creativity in diverse groups. In K. W. Phillips (Ed.), *Research on managing groups and teams* (Vol. 11, pp. 1-26). Bingley, UK: JAI. doi:10.1016/S1534-0856(08)11001-5
- Oakes, P., Turner, J., & Haslam, S. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology*, 30, 125-144.
- Oetzel, J. (2001). Self-construals, communication processes, and group outcomes in homogeneous and heterogeneous groups. *Small Group Research*, 32, 19-45. doi:10.1177/104649640103200102
- Oosterhof, A., van der Vegt, G., van de Vliert, E., Sanders, K., & Kiers, H. (2009). What's the difference? Insider perspectives on the importance, content, and meaning of interpersonal differences. *Journal of Occupational and Organizational Psychology*, 82, 617-637. doi:10.1348/096317908X342909
- Phillips, K., & Loyd, D. (2006). When surface and deep-level diversity collide: The effects on dissenting group members. *Organizational Behavior and Human Decision Processes*, 99, 143-160. doi:10.1016/j.obhdp.2005.12.001
- Phillips, K., Mannix, E., Neale, M., & Gruenfeld, D. (2004). Diverse groups and information sharing: The effects of congruent ties. *Journal of Experimental Social Psychology*, 40, 497-510. doi:10.1016/j.jesp.2003.10.003
- Phillips, K., Northcraft, G., & Neale, M. (2006). Surface-level diversity and decision-making in groups: When does deep-level similarity help? *Group processes & intergroup relations*, 9, 467-482. doi:10.1177/13684302060067557
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core team. (2009). nlme: Linear and nonlinear mixed effects models. R package version 3.1-93 [Computer software]. Retrieved from <http://cran.r-project.org/>
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437-448. doi:10.3102/10769986031004437
- R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software]. Retrieved from <http://www.R-project.org>



- Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen [QCM: A questionnaire to assess current motivation in learning situations]. *Diagnostica, 47*, 57-66. doi:10.1026//0012-1924.47.2.57
- Rico, R., Molleman, E., Sánchez-Manzanares, M., & van der Veegt, G. S. (2007). The effects of diversity faultlines and team task autonomy on decision quality and social integration. *Journal of Management, 33*, 111-132. doi:10.1177/0149206306295307
- Salas, E., Cooke, N., & Rosen, M. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors, 50*, 540-547. doi:10.1518/001872008X288457
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74.
- Schermuly, C. C., & Scholl, W. (IN PRESS). *Das Instrument zur Kodierung von Diskussionen (IKD)* [The discussion coding system (DCS)]. Göttingen, Germany: Hogrefe.
- Schippers, M. C., Hartog, D. N. D., Koopman, P. L., & Wienk, J. A. (2003). Diversity and team outcomes: The moderating effects of outcome interdependence and group longevity and the mediating effect of reflexivity. *Journal of Organizational Behavior, 24*, 779-802. doi:10.1002/job.220
- Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428. doi:10.1037/0033-2909.86.2.420
- Stegmann, S., & van Dick, R. (2009, May). *Does it matter what we think about diversity? A meta-analysis of the effects of diversity beliefs*. Paper presented at the 14th European Congress of Work and Organizational Psychology, Santiago de Compostela, Spain.
- Thatcher, S., Jehn, K., & Zanutto, E. (2003). Cracks in diversity research: The effects of diversity faultlines on conflict and performance. *Group Decision and Negotiation, 12*, 217-241. doi:10.1023/A:1023325406946
- Thomas, J., Bliese, P., & Jex, S. (2005). Interpersonal conflict and organizational commitment: Examining two levels of supervisory support as multilevel moderators. *Journal of Applied Social Psychology, 35*, 2375-2398. doi:10.1111/j.1559-1816.2005.tb02107.x
- Turner, J., Hogg, M., Oakes, P., Reicher, S., & Wetherell, M. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Basil Blackwell.
- van Dick, R., van Knippenberg, D., Hagele, S., Guillaume, Y., & Brodbeck, F. (2008). Group diversity and group identification: The moderating role of diversity beliefs. *Human Relations, 61*, 1463-1492. doi:10.1177/0018726708095711

- van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology, 89*, 1008-1022. doi:10.1037/0021-9010.89.6.1008
- van Knippenberg, D., Haslam, S., & Platow, M. (2007). Unity through diversity: Value-in-diversity beliefs, work group diversity, and group identification. *Group Dynamics: Theory, Research, and Practice, 11*, 207-222. doi:10.1037/1089-2699.11.3.207
- van Knippenberg, D., & Schippers, M. (2007). Work group diversity. *Annual Review of Psychology, 58*, 515-541. doi:10.1146/annurev.psych.58.110405.085546
- Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal, 36*, 590-602. doi:10.2307/256593
- Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. In R. Sutton & B. Straw (Eds.), *Research in organizational behavior* (Vol. 20, pp. 77-140). Greenwich, CT: JAI Press.
- Wittenbaum, G. M., & Stasser, G. (1996). Management of information in small groups. In J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition* (pp. 3-28). Thousand Oaks, CA: SAGE.
- Zellmer-Bruhn, M., Maloney, M., Bhappu, A., & Salvador, R. (2008). When and how do differences matter? An exploration of perceived similarity in teams. *Organizational Behavior and Human Decision Processes, 107*, 41-59. doi:10.1016/j.obhdp.2008.01.004

## Bios

**Bertolt Meyer** is a senior research and teaching associate in social psychology at the Department of Psychology at the University of Zurich, Zurich, Switzerland. His major research interests include team diversity, team performance, and leadership.

**Meir Shemla** is a postdoctoral fellow at the Department of Work and Organizational Psychology at Dresden University of Technology, Dresden, Germany. His major research interests include team diversity and leadership.

**Carsten C. Schermuly** is a postdoctoral fellow at the Department for Work, Organizational, and Social Psychology at Braunschweig University of Technology, Braunschweig, Germany. His major research interests include discussion coding, leadership, and psychological empowerment.