Article

Organizational Research Methods
1-42

DOI: 10.1177/1094428116639130
orm.sagepub.com
SAGE

Team Diversity and Categorization Salience: Capturing Diversity-Blind, Intergroup-Biased, and Multicultural Perceptions

Margarita Mayo[1], Daan van Knippenberg[2], Laura Guillén[3], and Shainaz Firfiray[4]

Abstract

It is increasingly recognized that team diversity with respect to various social categories (e.g., gender, race) does not automatically result in the cognitive activation of these categories (i.e., *categorization salience*), and that factors influencing this relationship are important for the effects of diversity. Thus, it is a methodological problem that no measurement technique is available to measure categorization salience in a way that efficiently applies to multiple dimensions of diversity in multiple combinations. Based on insights from artificial intelligence research, we propose a technique to capture the salience of different social categorizations in teams that does not prime the salience of these categories. We illustrate the importance of such measurement by showing how it may be used to distinguish among *diversity-blind* responses (low categorization salience), *multicultural* responses (positive responses to categorization salience), and *intergroup-biased* responses (negative responses to categorization salience) in a study of gender and race diversity and the gender by race faultline in 38 manufacturing teams comprising 239 members.

Keywords

multivariate analysis, computational modeling, team diversity, categorization salience, leadership

[1]IE Business School, Madrid, Spain
[2]Erasmus University Rotterdam, Rotterdam, Netherlands
[3]ESMT European School of Management and Technology, Berlin, Germany
[4]Warwick Business School, University of Warwick, Coventry, UK

Corresponding Author:
Margarita Mayo, IE Business School, Alvarez de Baena 4, Madrid 28006, Spain.
Email: margarita.mayo@ie.edu

Diversity has become a fact of life for many organizations. The changing nature of the workforce and the popularity of work teams are bringing more people to work with others who differ in their demographic backgrounds (Jackson, Joshi, & Erhardt, 2003; Joshi & Roh, 2010). This increased diversity is one of the main challenges of today's organizations (Harrison, Price, Gavin, & Florey, 2002). Demographic diversity can be a positive influence when associated differences in perspectives stimulate a more in-depth understanding of the issues at hand, and better quality and more innovative solutions to problems and decisions (van Knippenberg, De Dreu, & Homan, 2004; cf. diversity as variety; Harrison & Klein, 2007). At the same time, demographic diversity may invite interpersonal tensions and factional thinking within a team that inhibit the stimulating influence of diversity and are detrimental to team performance (van Knippenberg et al., 2004; cf. diversity as separation; Harrison & Klein, 2007). An important challenge thus is to understand the factors that prevent the negative effects of team diversity and stimulate its positive effects—a challenge that puts a premium on an understanding of the contingencies of diversity effects (van Knippenberg & Schippers, 2007).

The potential of demographic diversity to have negative performance effects has traditionally been viewed through the lens of the social categorization perspective (van Knippenberg & Schippers, 2007; Williams & O'Reilly, 1998). This perspective has gone through a development that we aim to take one important step further in the present study—a development that requires a new measurement technique to adequately and efficiently capture the salience (i.e., cognitive activation) of categorizations (i.e., "us versus them" perceptions) based on multiple demographic attributes. Initially, the social categorization perspective posited that demographic diversity invites social categorization based on demographic differences that result in intergroup biases—negative responses to demographically different others who are seen as not part of one's in-group—that disrupt team process and performance (Williams & O'Reilly, 1998). This perspective, however, would imply a negative main effect of demographic diversity that is not supported by the evidence (van Dijk, van Engen, & van Knippenberg, 2012). An important step in the development of the social categorization perspective has been the emphasis on *categorization salience*—the extent to which a social categorization is cognitively activated and is used by group members as a basis for distinguishing in-groups and out-groups (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987; van Knippenberg et al., 2004). This salience perspective recognizes that demographic differences do not necessarily invite social categorization and that contextual factors that foster or hinder the diversity-salience relationship need to be taken into account in understanding diversity effects.

Thus, understanding which social categories are more important in a specific group is a critical research question for diversity scholars. However, we know very little about how the context makes some social identities more psychologically prominent in the minds of team members than others. Categorization salience is more often evoked than actually measured. Moreover, diversity scholars interested in testing categorization salience models face problems when it comes to measuring the salience of multiple dimensions of diversity in combination. The purpose of our article is to introduce a measure of categorization salience that addresses problems with earlier measures and allows researchers to test more complete models of how contextual factors influence the effects of diversity paving the way for better diversity management practices. A starting point for the current contribution is the observation that categorization salience in diversity research (e.g., Pearsall, Ellis, & Evans, 2008; Randel, 2002; Randel & Jaussi, 2003) assumes that once a categorization is salient, it invites disruptive effects. However, it is increasingly recognized that salient categorizations may also invite positive responses (Ely & Thomas, 2001; van Knippenberg, van Ginkel, & Homan, 2013). An important methodological implication of these insights is that diversity research cannot suffice with measuring potential precursors to categorization salience like diversity on single attributes and their combination (i.e., where negative effects of diversity are assumed to reflect salient categorizations, and the absence of negative effects signal low categorization salience). Rather, diversity research needs to directly assess categorization salience in diverse teams to distinguish what we call *diversity-blind* responses (low categorization salience), *multicultural* responses (positive responses to salient categorization), and *intergroup-biased* responses (negative responses to salient categorization). Existing measures are not suitable for effectively assessing the salience of multiple alternative categorizations (e.g., Earley & Mosakowski, 2000; Randel, 2002). Thus, the key contribution of the current study is that we develop a measure of categorization salience based on insights from artificial intelligence (ID3; Quinlan, 1986) that addresses these problems with earlier measurement.

We also illustrate the importance of measuring categorization salience directly by showing how gender and race diversity and their combination in gender by race faultlines can invite diversity-blind, multicultural, and intergroup-biased responses. This illustration does not only help establish the usefulness of our proposed measure, but constitutes a secondary contribution of our study as being the first to explicitly distinguish diversity-blind, multicultural, and intergroup-biased responses empirically. The current study provides the methodological tools as well as the conceptual insights to advance our understanding of the effects of team diversity in terms of the important role of categorization salience.

Theoretical Background

Team Diversity and Categorization Salience

Team diversity refers to the degree to which team members differ on any given attribute (van Knippenberg & Schippers, 2007) and thus in principle could concern any attribute on which people may differ—demographic attributes, functional/educational background, or psychological traits or states. In practice, diversity research concerns itself primarily with a more limited set of demographic and functional/educational attributes and diversity management in practice tends to by and large concern the "Big Two" of diversity management—gender and race/ethnicity/cultural background (cf. van Knippenberg, Homan, & van Ginkel, 2013). Most of the issues with negative effects associated with diversity—stereotyping, discrimination, and disrupted performance—are also associated with these "Big Two" of gender and race (van Dijk et al., 2012). A focus on gender diversity and race diversity thus seems particularly relevant in speaking to the performance effects of diversity. This focus is also supported by research on social categorization, which showed that individuals spontaneously categorize others using a superordinate category that represented sex and race simultaneously because it increases "informativeness" (Stangor, Lynch, Duan, & Glass, 1992).

In understanding the team performance effects of gender and race diversity, the primary conceptual basis in diversity research was and still is the social categorization perspective. Earlier readings of this perspective seemed to imply more or less that gender or race differences between team members would inevitably invite social categorization in terms of own group and other group (categorization salience), and such categorization would equally inevitably invite intergroup biases that would disrupt performance by reducing the willingness to collaborate with demographically different team members (Williams & O'Reilly, 1998). A more recent reading of this perspective that is reflected in the *categorization-elaboration model* (CEM) of team diversity and performance (van Knippenberg et al., 2004), however, recognizes that neither the path from diversity to categorization salience nor the path from categorization salience to intergroup bias inevitably occurs but rather is contingent on other factors. This is an important observation because it points to diversity management opportunities either in preventing diversity from inviting categorization salience or in preventing categorization salience from resulting intergroup biases.

The important implication of this analysis is that from a social categorization perspective it is meaningful to distinguish between three types of responses to gender and race diversity. The first is

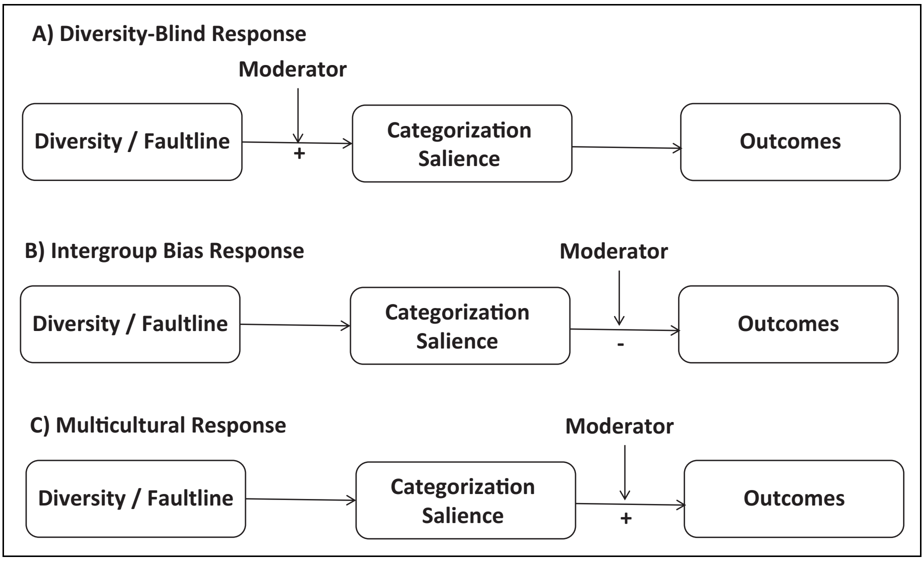


Figure 1. Three responses to diversity.

what can be called a *diversity-blind* response (cf. Plaut, Thomas, & Goren, 2009); gender or race differences, or their combination gender-by-race, do not invite the cognitive activation of gender or race categories. The second is an *intergroup-biased* response; gender and/or race differences invite salient categorization resulting in intergroup biases. The third is a *multicultural* response (cf. Plaut et al., 2009); gender and/or race differences invite salient categorization, which is embraced as a positive characteristic of the team (see Figure 1).

The distinction between diversity-blind, intergroup-biased, and multicultural responses is important because it implies different team performance outcomes. The CEM posits that diversity does not only have the potential to invite intergroup biases that disrupt performance, but also represents a source of diverse insights, information, and perspectives that can benefit performance. Benefiting from diversity as an informational resource implies actively engaging with diversity, however—seeking out differences to benefit performance in a process of information exchange and integration. Preventing intergroup bias is important here because biases disrupt performance and thus also stand in the way of benefiting from diversity as an informational resource. Preventing intergroup bias should not be equated with stimulating the use of diversity as an informational resource, however; the use of diversity as an informational resource requires motivated, proactive engagement with diversity and not merely the absence of intergroup bias (Kearney, Gebert, & Voelpel, 2009; Nederveen Pieterse, van Knippenberg, & van Dierendonck, 2013; cf. van Ginkel & van Knippenberg, 2008). Whereas intergroup-biased responses to diversity thus are clearly the least preferred outcome in that they should result in negative performance effects of diversity, diversity-blind and multicultural responses to diversity should have other effects (cf. Plaut et al., 2009). Diversity-blind responses preempt intergroup bias by suppressing social categorization, but by the same token they discourage proactive engagement with diversity as an informational resource. Such responses should thus be associated with the absence of diversity effects on performance rather than with positive effects. Multicultural responses in contrast are associated with a positive recognition of different social categories and should thus be most likely associated with positive diversity effects on performance.

In reference to categorization salience, it is important to note that we conceptualize and operationalize salience as a team level construct—*team salience*. Past research in categorization has treated salience as an individual-level construct for understanding individual perceptions (e.g., Randel, 2002). For the team level analysis of the diversity-performance relationship, however, we suggest that members of a group categorize themselves and others on the basis of patterns of differences on certain attributes. Team members tend to agree on the nature of those differences; yet how they experience those differences and their consequences vary across team members. For example, consider a team of three men and three women faculty members of an Organizational Behavior department. At the team level, members of this team seem to clearly form subgroups on the basis of gender differences. During lunch, men sit together with other men and women with other women; on the weekends men get together for golf and women meet to play tennis. All men and women would agree that gender matters when describing and thinking about their work team. Regardless of how important gender is in their personal identity, at the team level, gender explains in part the social dynamics within their team. Thus, team salience is a complex function that provides a better understanding of team member arrangements on the basis of the salience of certain demographic categories.

A given social category (e.g., sex) is collectively salient or psychologically important in the minds of team members *only* when it is widely used by group members to categorize themselves into in-groups and out-groups. The approach we propose deals with team salience based on in-group versus out-group categorizations among all possible dyadic relationships within the team. When people in the team form subgroups, it is rather clear to other team members who is part of the in-group and the out-group and which dimensions they follow to divide the team. For example, imagine that we are observing a Sales and Marketing group comprising five members formed by three women and two men, all of them with a management background, and we want to predict the extent to which gender might help to explain who might sit together for lunch on a daily basis. If we asked them to make in-group versus out-group categorizations with every other member of the team, it would reveal that the three women rate other women as in-group and men as out-group. Likewise, the two men perceive themselves as belonging to the same subgroup but see women as the out-group. In this case, it is clear that the attribute of sex serves to explain the in- versus out-group social categorizations within the team and it is highly salient in the mind of all team members. Now consider a Research & Development group also comprising three women and two men. Two of the women and one man have an engineering background whereas one woman and one man have a management background. When we asked them to make in- versus out-group judgments, the three engineers give in-group ratings to one another and see their two management colleagues as out-group. In this case, the attribute of educational background, not sex, is the basis for their in-group/out-group categorizations.

The Need for a Categorization Salience Measure

While in earlier understandings of social categorization processes in diverse teams negative effects of diversity were seen as the inevitable effect of salient categorizations, the observation that diversity had a negative effect could be presumed to imply categorization salience, whereas the absence of a negative effect would imply low salience (cf. van Knippenberg et al., 2004). Thus, to further develop diversity theory and provide stronger tests of theoretical perspectives, assessment of categorization salience is necessary.

In this respect, diversity research encounters a methodological challenge. There have been relatively few attempts to measure categorization salience, and the measures that have been used are less suitable to research on multiple dimensions of diversity. Both Earley and Mosakowski (2000) and Randel (2002) assessed categorization salience through a questionnaire measuring self-reported cognitive activation of a given category (nationality, sex). For example, Randel (2002, p. 750) defines salience as “an individual-level measure of how prominently a demographic category is used to describe one's work group members (for example, ''I think of my work group in terms of women and men''). Individual members needed to complete a three-item scale to assess how important gender was to describe their team. A sample item was ''when people ask me about who is in the group, I initially think of describing group members in terms of gender composition (e.g., two women and three men).'' Yet, one of the weaknesses of Randel's (2002) salience measure is that it may prime group members to consider the salience of gender even if they were not thinking of this social category in the first place. We would not argue with the use of such a measure for the research purposes of Earley and Mosakowski (2000) or Randel (2002), which were both studies zooming in on a single diversity attribute. As soon as one is interested in multiple diversity attributes, however, and especially when one is also interested in their combination, measures like those employed by Earley and Mosakowski (2000) and Randel (2002) become problematic, because they require that the categorization of interest be explicitly identified.

Importantly, research on diversity faultlines suggests that assessing the salience of multiple categorizations is exactly what should be on the agenda for diversity research. Lau and Murnighan (1998) propose the concept of *faultline*, which refers to ''hypothetical dividing lines that may split a group into subgroups based on one or more attributes'' (p. 328). They argue that diversity in work teams is very complex and does not comprise single attributes, but rather of multiple demographic attributes that subdivide groups. Groups that have identical dispersion of demographic attributes can still have different dynamics if those characteristics are aligned among the individuals in the group so that when the group is divided on the basis of one attribute (e.g., sex), then these subgroups are very similar with respect to another attribute (e.g., race). For example, in a group of five, all three females are White and all two males are Black. When multiple demographic characteristics correlate with one another, divisions in groups occur on the basis of faultlines (Lau & Murnighan, 1998). Existing research on social categorization shows that people use more than one attribute to categorize others, and thus single demographic attributes may provide an insufficient measure of diversity in groups (Thatcher, Jehn, & Zanutto, 2003). When multiple demographic characteristics correlate with one another, divisions in groups will occur on the basis of faultlines. Faultline measures are ''raw data'' determinants of psychological salience, but not salience itself. It is therefore important to distinguish between faultlines and categorization salience, or between ''faultlines and the earthquakes the faultlines may cause'' (Shaw, 2004, p. 69). Thus, the existence of a faultline in the group does not necessarily mean the faultline erupts in the minds of the people to produce real *earthquakes* that disrupt team performance. Accordingly, a measurement of social categorization salience is needed to account for multiple diversity attributes and their faultlines.

This implies that to assess the salience of, for instance race and gender categorizations as well as a categorization based on their combination (e.g., Black women), one would need to administer a questionnaire similar to Randel's three times—for sex, for race, and for their combination. Once one is interested in more than one dimension and their faultlines, things get exponentially more difficult. Studying for instance sex, race, and functional background diversity and their faultlines would require administering the same questionnaire seven times—three times to capture the salience of categorizations based on each diversity attribute by itself, three times to capture the salience of categorizations based on the combination of two of the attributes, and once to capture the salience of the categorization based on all three attributes in combination (cf. van Knippenberg, Dawson, West, & Homan, 2011). A first problem here is that the repeated use of the same scale becomes unwieldy in terms of survey length and respondent annoyance.

This in and of itself is enough to argue for a salience measure that does not assess salience for each categorization separately in a survey. There are two additional considerations to be considered. First, for the salience of faultlines, items in a survey format become very complicated—even the simple division between Black men, Black women, White men, and White women arguably asks for items that refer either to all these combinations or refer to the race-sex coupling in a more abstracted sense. For faultlines involving more than two diversity attributes, such items might be beyond comprehension for a number of respondents. Second, for faultlines and single-attribute diversity asking directly about salience could actually prime respondents to make categorizations more salient even if they were not thinking of these categories previously—the measurement itself may influence responses.[1] In sum, the state of the science in diversity research puts a premium on a new way of measuring categorization salience.

Measuring Categorization Salience

To address the shortcomings of current survey measures of social categorization salience in diverse teams, we propose an artificial intelligence framework to operationalize the salience of social categories. For the sake of simplicity, we will limit the presentation here to the "Big Two" diversity categories of sex and race and the faultline formed by sex-by-race combinations, but our proposed approach can easily be extended to include more diversity attributes and faultlines.

More specifically, we use the leading machine learning system named the Interactive Dichotomizer 3 (ID3; Quinlan, 1983) because of its structural and functional similarities to human categorization processes. The ID3 not only classifies the dependent variables as does for example discriminant analysis, but also creates an entire classification structure represented in a rule decision tree in which each attribute is associated with a number representing the amount of information that that particular attribute provides to the entire system. The ID3 learning system provides a useful model for understanding social categorization for several reasons. First, the ID3 has a functional similarity to social categorization. As individuals classify people into categories on the basis of their relative similarity, the ID3 is particularly designed to deal with classification problems. Second, the ID3 parallels social categorization in that both use categories to represent knowledge. As the ID3 organizes the acquired knowledge of the system around attributes or nodes of a decision tree, individuals also organize their social knowledge in the form of social attributes. Third, the implicit rule used by individuals to rank these social attributes from the most salient to the least salient is an information rule. The most informative attribute to categorize other team members into in-group and out-group is the one most salient in our mind (Fiske & Taylor, 1991). Similarly, the ID3 uses an information-based rule to rank attributes, placing the most informative attribute at the top of the decision tree. Fourth, social categorization and ID3 are comparable with respect to the underlying learning strategy. Both knowledge structures are acquired through an inductive learning strategy. The ID3 parallels the process by which social categories are formed in that the formation of these social categories is the result of an *inductive* process by which, over the years, people extract knowledge from specific examples that guide the development of the category (Fiske & Taylor, 1991).

The Process of Measuring Category Salience

In applying the ID3 algorithm, four phases need to be followed. In Phase 1, a social categorization variable (in-group vs. out-group) has to be computed. In Phase 2, the diversity variables need to be selected. Phase 3 requires the creation of dummy diversity variables at the dyadic level (relational diversity). Finally, Phase 4 requires running the ID3 algorithm to assess how salient the diversity attributes and their combinations are within a group. Let us consider a work team from our sample (Team 1) comprising 6 people as an example. The demographic profile of this team is as follows: Person 1 = White female, Person 2 = White male, Person 3 = Black male, Person 4 = White female, Person 5 = White female, and Person 6 = White male (see Table 1).

Table 1. Example of a Work Team From the Data Set.

Person 1	Person 2	Person 3	Person 4	Person 5	Person 6
White Male	White Male	Black Male	White Female	White Female	White Male

Dyad	Relational Sex	Relational Race	Social Categorization
Person 1 → Person 2	0	0	0
Person 1 → Person 3	0	1	0
Person 1 → Person 4	1	0	0
Person 1 → Person 5	1	0	1
Person 1 → Person 6	0	0	1
Person 2 → Person 1	0	0	0
Person 2 → Person 3	0	1	1
Person 2 → Person 4	1	0	0
Person 2 → Person 5	1	0	1
Person 2 → Person 6	0	0	1
Person 3 → Person 1	0	1	0
Person 3 → Person 2	0	1	0
Person 3 → Person 4	1	1	0
Person 3 → Person 5	1	1	0
Person 3 → Person 6	0	1	0
Person 4 → Person 1	1	0	0
Person 4 → Person 2	1	0	0
Person 4 → Person 3	1	1	0
Person 4 → Person 5	0	0	1
Person 4 → Person 6	1	0	0
Person 5 → Person 1	1	0	0
Person 5 → Person 2	1	0	1
Person 5 → Person 3	1	1	0
Person 5 → Person 4	0	0	0
Person 5 → Person 6	1	0	1
Person 6 → Person 1	0	0	1
Person 6 → Person 2	0	0	1
Person 6 → Person 3	0	1	1
Person 6 → Person 4	1	0	0
Person 6 → Person 5	1	0	1

Phase 1: Computing the Social Categorization Variable

Literature on social categorization processes suggests that categorization takes a dichotomous form as people naturally divide others by a process of in-group versus out-group classifications (cf. Crocker, Fiske, & Taylor, 1984; Rosch, 1975; Shaw, 2004; Turner, 1987). According to Turner (1987), "Category formation is relative to the frame of reference (the pool of psychologically relevant stimuli) and hence the available contrasts provided by the salient stimulus field, and depends not just on 'similarities' between stimuli, as is usually assumed, but on *relative* similarities, on *more* similarity (or less difference) between certain stimuli than between those and others" (p. 47).

To apply the measure of categorization salience that we are proposing, participants should therefore be given in a survey the names of all members of their team and asked to categorize them as belonging to in-groups versus out-groups. Tajfel and colleagues (e.g., Billig & Tajfel, 1973) noted that the variable of social categorization always coincides with the variable of similarity. It does not matter how the subjects divided themselves into subgroups, the important point is that the subjects always perceived similarity between their fellow in-group members and themselves, and dissimilarity between themselves and the members of the out-group. Thus, it is reasonable to assume that similarity ratings can be used as a proxy of social categorization. In this example, we chose to use a continuous similarity variable and then dichotomize it. Participants rated how (dis)similar each of them is to oneself, using a scale ranging from 1 (*very dissimilar*) to 10 (*very similar*). This continuous variable was then coded as a dichotomous social categorization variable with two values (see Table 1 column social categorization): in-group (1; perceived similarity scores greater than or equal to 5), and out-group (0; perceived similarity scores less than 5).[2] Regardless of whether researchers use direct categorization measures (i.e., asking group members to make in- vs. out-group judgments), or similarity ratings, what is important is that the ID3 explores categorization as a categorical variable.

Phase 2: Selection of Diversity Attributes

To be able to complete the data set for each team, we needed to decide on the diversity attributes that might impact their in-group versus out-group categorizations, like, for example, demographic attributes (e.g., sex, age, race), personality (e.g., extroversion, neuroticism), and/or values (e.g., individualism, collectivism). It was, however, beyond the scope of this article to attempt to provide a comprehensive list of attributes to study diversity in groups. Since we were interested in illustrating the development of a salience measure for each diversity attribute chosen, the exact nature of the attributes involved was a theoretical issue, and, for the sake of simplicity, we limited the presentation here to the "Big Two" diversity categories of sex and race and the faultline formed by sex-by-race combinations.

Phase 3: Creation of Dummy Relational Diversity Variables (Dyadic Level)

The next step is to code the diversity variables in an interval scale and code relational scores (See Table 1 columns Relational Sex and Relational Race). Decision tree algorithms (such as the ID3) tend to perform better when dealing with discrete, categorical features (Kotsiantis, 2007). Drawing from the diversity literature, there are variables that can thus be straightforwardly used with the ID3 approach, which include, among others, race, sex, occupational level, organizational role, education, career level, legal marital status, and children. Although the ID3 algorithm does not directly deal with continuous predictors, we discuss this point further in the conclusion. For this study, we coded our two demographic variables as follows: gender (two levels coded $male = 0, female = 1$) and race (four levels coded $White = 1$, $Black = 2$, $Hispanic = 3$, $Asian = 4$). Then, we created the relational variables based on dyadic comparisons among all members of the team with a value of "0" indicating that two group members share the attributes (e.g., same sex, or same race), and a value of "1" indicating that they do not share the attribute.

Phase 4: Running the ID3 Algorithm

The last step is to run the ID3 algorithm for each of the groups in the data set to obtain the relative salience of each diversity attribute (e.g., sex and race) and their faultlines (e.g., sex-by-race faultline) in each work group. The ID3 algorithm determines the extent to which the selected diversity attributes and their combinations predict social categorization. As mentioned earlier, a given diversity attribute (e.g., sex or race) is supposed to be salient in the minds of the team members when they use it as the basis for social categorization judgments into similar (in-group members) or different (out-group members) from oneself. Following the recommendations of Thompson and Thompson

(1986), we show how the ID3 algorithm determines the amount of information contained in each attribute when classifying group members into in-group and out-group drawing from separate datasets for each work team. The advantage of this approach is that regardless of the number of diversity attributes, and their combinations, the task for respondents is the same—rating perceived similarity between themselves and each of their fellow team members. Thus, participants are not primed by the question to consider the salience of particular social categories, even if they were not thinking of these categories previously.

A One-Team Illustration

We illustrate the procedure of the ID3 algorithm (Quinlan, 1986) to measure salience of two social categories (sex and race) within a team. We show how the ID3 algorithm determines the amount of information contained in each attribute when classifying group members into in-group and out-group considering our previous example of the work team in Table 1.

We first create a data set for the team that includes 30 dyadic cases, 12 of which are in-group and 18 out-group categorizations. The *social categorization* column refers to the in-group and out-group variable and it is calculated based on group members' ratings of each other as similar to oneself (in-group) or different from oneself (out-group). The *relational sex* and *relational race* columns refer to relational demography on these attributes for all dyads within the team. These relational demography variables can have two values—same or different—and are used to predict the classification of these 30 cases into in-groups and out-groups.

Because some attributes give more information about how to classify a group member into an in-group or out-group than other attributes do, it is possible to calculate the amount of information gained by considering each attribute separately. The ID3 algorithm uses the notion of entropy as defined in information theory (Pierce, 1980). The concept of *entropy* applied to social categorization refers to the uncertainty of the classification of an individual into an in-group or out-group and it ranges from 0 to 1. The entropy is 0 if all cases belong to the same class because the data are perfectly classified (e.g., all cases are out-groups). In contrast, the entropy is 1 if the classification of the cases is totally random (50% in-groups and 50% out-groups). The purpose of the ID3 is to determine the gain in information provided by each attribute (sex and race) when deciding whether the social categorization is in-group or out-group. The attribute that provides the highest gain is used as the root node in the decision tree and we take it as the most salient attribute in the mind of team members. Mathematically, if an individual can be classified into N different classes, $c_1, \ldots c_N$, and the probability of an individual being in class i is $p(c_i)$, then the entropy of classification, H (C), is as follows:

Equation 1:

$$\mathrm{H}\,(\mathrm{C}) = -\sum_{i=1}^{N} \mathrm{p}(c_i)\, \log_2 \mathrm{p}(\mathrm{c_i})\ \mathrm{H}\,(\mathrm{C}) = -\sum_{i=1}^{N} \mathrm{p}(c_i)\, \log_2 \mathrm{p}(\mathrm{c_i})$$

where $p(c_i)$ is the proportion of individuals who belong to each class of the social categorization variable and the log base 2 represents the number of bits needed to represent that many different individuals.[3]

Applying the above equation to the example set in Table 1 gives us a measure of uncertainty (or entropy) about social categorization being in-group or out-group. Because there are two possible values for social categorization (in-group and out-group), the probability of having an in-group value is 12 out of a total sample set of 30. The probability of out-group is 18/30. Thus, the entropy of classification for the total data set is

$$\mathrm{H}\,(\mathrm{C}) = -\,\mathrm{p}(\mathrm{in-group})\, \log_2 \mathrm{p}(\mathrm{in-group}) - \mathrm{p}(\mathrm{out-group})\, \log_2 \mathrm{p}(\mathrm{out-group})$$

$$-(12/30)\ \log_2(12/30) - (18/30)\ \log_2(18/30) = 0.97\text{bits}$$

This value is the total degree of entropy in the data set. The second step is to determine the amount of information contained in each social category (sex and race). To calculate the entropy of classification after deciding on a particular category, represented by H(C/A), the ID3 algorithm splits the example set into subsets where each case has the same value of the partitioning category (i.e., same sex or different sex).

Salience of Sex Identity. We calculate information gained by considering the *sex* category. After splitting by sex, the *same sex branch* gives a total of 14 cases including 7 in-group cases and 7 out-group cases, and the *different sex branch* gives a total of 16 cases including 5 in-group cases and 11 out-group cases. Let us calculate the entropy for each sex branch.

- Entropy for "same" sex branch = – (7/14) $\log_2$ (7/14) – (7/14) $\log_2$ (7/14) = $\underline{1}$ bit
- Entropy for "different" sex branch = – (5/16) $\log_2$ (5/16) – (11/16) $\log_2$ (11/16) = $\underline{0.90}$ bits

To calculate the entropy of the entire data set split on "sex," represented as H(C/sex), the ID3 method computes the weighted average of the entropy of each subset (14 *same sex* and 16 *different sex*). In the present example, this yields the following:

- Entropy (System, sex) = (14/30) (1) + (16/30) (.90) = $\underline{0.94}$ bits

The information gained by considering sex is the total degree of entropy in the data set minus the entropy of the system when we consider sex. We take this value as indication of the degree of salience of the sex category. Thus, the salience of sex identity in this team is:

$$\text{Entropy (S)} - \text{Entropy (S, sex)} = 0.97 - 0.94 = 0.03 \text{ bits}$$

Salience of Race Identity. We repeat the same procedure to calculate the information gained by considering the *race* category. After splitting by race, the *same race branch* gives a total of 20 cases including 10 in-group cases and 10 out-group cases, and the *different race branch* gives 10 cases including 2 in-group cases and 8 out-group cases. Let us calculate the entropy for each race branch.

- Entropy for "same" race branch = – (10/20) $\log_2$ (10/20) – (10/20) $\log_2$ (10/20) = $\underline{1}$ bit
- Entropy for "different" race branch = – (2/10) $\log_2$ (2/10) – (8/10) $\log_2$ (8/10) = $\underline{0.72}$ bits

$$\text{Entropy (S, race)} = (20/30)\ (1) + (10/30)\ (0.72) = 0.91\text{bits}$$

Information gained by testing race is:

$$\text{Entropy (S)} - \text{Entropy (S, race)} = 0.97 - 0.91 = 0.06 \text{ bits}$$

Race has the highest gain; therefore, the race category is most salient in the mind of team members because it provides the highest amount of information when categorizing others into in-group and out-group (salience of race = 0.06, and salience of sex = 0.03). The amount of information associated with each social category is taken as an indication of the degree of collective salience of each social category within the team. Since the entropy of the classification into in-groups and out-groups after partitioning by race gives the smallest entropy or, equivalent, the highest information, the category "race," is chosen to serve as the root node of the decision tree. When "race" is the starting point of the tree, the tree takes the form shown in Figure 2. This figure also

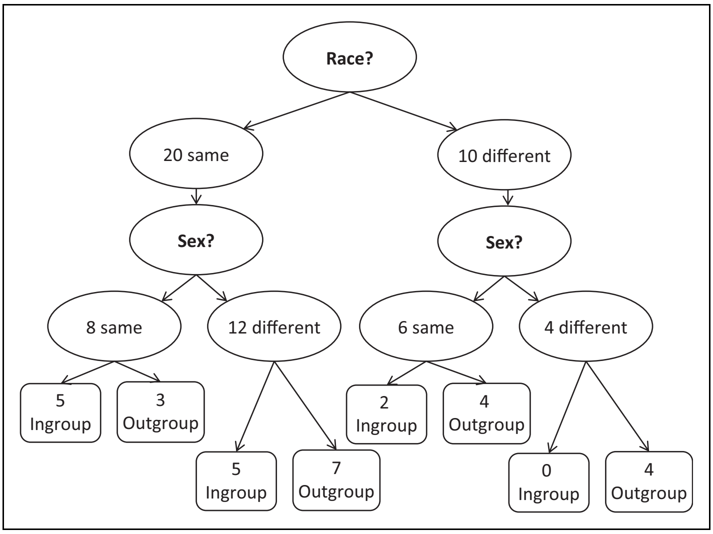


Figure 2. Decision tree of the empirical illustration.

shows a further split based on the category "sex" and the value of the social categorization into in-groups and out-groups.

Salience of Sex-by-Race Faultline. So far, the analysis is only concerned with salience of dimensions of diversity in isolation (i.e., sex salience and race salience). However, it is also possible to compute the interaction between different dimensions (i.e., sex-by-race) to assess the salience of a single superordinate identity that represents both dimensions simultaneously. Following our example, we present next how to capture the salience of race and sex in combination. Using "race" as the root node, we once again applied the ID3 algorithm to each of the two branches considering the two levels of the sex dimension. Step 1 describes the procedure for the "same race" branch and Step 2 describes the procedure for the "different race" branch.

Step 1. When we consider the "same race" branch, there still remains maximum entropy (1 bit) because half of the cases are in-group and half are out-group (see Table 2a). Then, we applied the algorithm for the two values of the sex dimension (i.e., same or different) to assess the amount of information gained when also considering sex. After further splitting by sex, the *same sex branch (of the same race)* gives a total of eight cases including five in-groups and three out-groups and the *different sex branch* (*of the same race)* gives a total of twelve cases including five in-groups and seven out-groups. Let us calculate the entropy for each sex branch given same race.

Step 1: Entropy of "same race" branch before considering for sex = – (10/20) $\log_2$ (10/20) – (10/20) $\log_2$ (10/20) = $\underline{1}$ bit

- Entropy for same sex branch = – (5/8) $\log_2$ (5/8) – (3/8) $\log_2$ (3/8) = $\underline{0.95}$ bits
- Entropy for different sex branch = – (5/12) $\log_2$ (5/12) – (7/12) $\log_2$ (7/12) = $\underline{0.98}$ bits

$$\text{Entropy}\ (S_{\text{same race}},\ \text{sex}) = (8/20)\ (0.95) + (12/20)\ (0.98) = \underline{0.97}\text{bits}$$

Step 2. Next when we consider the "different race" branch (see Table 2b), there still remains entropy of 0.72 bits. Then, we applied the algorithm for the two categories of the sex dimension

Table 2a. Data Subset for "Same Race" Branch Testing for Sex.

Dyad	Relational Sex	Relational Race	Social Categorization
Person 1 → Person 2	0	0	0
Person 1 → Person 4	1	0	0
Person 1 → Person 5	1	0	1
Person 1 → Person 6	0	0	1
Person 2 → Person 1	0	0	0
Person 2 → Person 4	1	0	0
Person 2 → Person 5	1	0	1
Person 2 → Person 6	0	0	1
Person 4 → Person 1	1	0	0
Person 4 → Person 2	1	0	0
Person 4 → Person 5	0	0	1
Person 4 → Person 6	1	0	0
Person 5 → Person 1	1	0	0
Person 5 → Person 2	1	0	1
Person 5 → Person 4	0	0	0
Person 5 → Person 6	1	0	1
Person 6 → Person 1	0	0	1
Person 6 → Person 2	0	0	1
Person 6 → Person 4	1	0	0
Person 6 → Person 5	1	0	1

Table 2b. Data Subset for "Different Race" Branch Testing for Sex.

Dyad	Relational Sex	Relational Race	Social Categorization
Person 1 → Person 3	0	1	0
Person 2 → Person 3	0	1	1
Person 3 → Person 1	0	1	0
Person 3 → Person 2	0	1	0
Person 3 → Person 4	1	1	0
Person 3 → Person 5	1	1	0
Person 3 → Person 6	0	1	0
Person 4 → Person 3	1	1	0
Person 5 → Person 3	1	1	0
Person 6 → Person 3	0	1	1

(i.e., same or different) to assess the amount of information gained when testing for sex. After further splitting by sex, the *same sex branch (of the different race)* gives a total of six cases including two in-groups and four out-groups and the *different sex branch* (*of the different race)* gives a total of four cases including all out-groups. Let us calculate the entropy for each sex branch given different race.

Step 2: Entropy of "different race" branch before considering for sex = – (2/10) $\log_2$ (2/10) – (8/10) $\log_2$ (8/10) = 0.72 bits

- Entropy for same sex branch = – (2/6) $\log_2$ (2/6) – (4/6) $\log_2$ (4/6) = 0.92 bits
- Entropy for different sex branch = 0 (all out-group cases)

$$\text{Entropy}\ (S_{\text{different race}},\ \text{sex}) = (6/10)\ (.92) + (4/10)\ (0) = \ 0.55\ \text{bits}$$

To calculate the entropy of the entire data set considering race and sex in combination, we next compute the weighted average of the entropy of each subset of race testing for sex (twenty cases of same race and ten cases of different race).

$$\text{Entropy (S, race} \times \text{sex)} = (20/30)\ (0.97) + (10/30)\ (0.55) = 0.83 \text{ bits}$$

Information gained by considering sex $\times$ race is:

$$\text{Entropy (S)} - \text{Entropy (S, sex} \times \text{race)} = 0.97 - 0.83 = 0.14 \text{ bits}$$

Salience of sex $\times$ race would take the value of 0.14, meaning that in combination the sex and race categories are used by team members to divide themselves into in-groups and out-groups to a greater extent than the sex or the race category in isolation (0.04 and 0.06 levels of salience, respectively).

To simplify the application of the ID3 salience measure, we have developed an Excel macro. Table 3 illustrates the application of this ID3 Excel macro for the above example in a step-by-step way. Readers interested in applying the ID3 algorithm can find an EXCEL macro in the link provided as a endnote.[4]

An Empirical Illustration: Diversity and Leadership

We further illustrate the use of our proposed technique as well as its usefulness in diversity research in a survey of work teams. Again, we focus on the "Big Two" sex and race diversity attributes and the sex-by-race faultline. Because the evidence in diversity research overwhelmingly suggests that diversity attributes may have positive, negative, or null effects contingent on moderators (van Dijk et al., 2012; van Knippenberg & Schippers, 2007), we also include a focus on moderation in this illustration. In particular, we focus on charismatic-transformational leadership because it is the only moderating influence that has been studied both for the effects of single-attribute diversity (Greer, Homan, De Hoogh, & Den Hartog, 2012; Kearney & Gebert, 2009; Shin, Kim, Lee, & Bian, 2012; Shin & Zhou, 2007) and diversity faultlines (Kunze & Bruch, 2010), with diversity effects and faultline effects more positive/less negative in the presence of higher, rather than lower, charismatic-transformational leadership. A focus on the moderating role of charismatic-transformational leadership (for short: charismatic leadership) is thus well-suited, because it speaks to the only common denominator in research on the moderators of the effects of diversity and diversity faultlines. Even so, the focus on charismatic leadership comes with an important caveat. As van Knippenberg and Sitkin (2013) outline, the concept of charismatic leadership is problematic both in its poor conceptualization and in terms of its poor measurement—the most plausible reading of the evidence is that survey measures of charismatic leadership reflect the subjective perception of effective leadership. We would therefore not argue for any conceptual contribution flowing from its study. For the present purposes, however—the illustration of our approach to the measurement of categorization salience—the focus on charismatic leadership is useful to illustrate what better measurement of categorization salience can bring to the diversity field.

The social categorization analysis suggests that both sex diversity and race diversity and sex-by-race faultlines can render social categorizations based on these attributes' salience. Neither diversity nor diversity faultlines deterministically lead to the salience of the associated categorization—team members can be diversity-blind—and categorization salience cannot be equated with negative effects—responses can be intergroup-biased but also multicultural. It is therefore important to assess both the relationship between diversity and salience, and the mediating role of salience in the relationship between diversity and relevant outcomes. For our illustration, it is thus important to assess the relationship between sex diversity and sex categorization salience, between race diversity and race categorization salience, and between the sex by race faultline and the salience of the categorization capturing both sex and race. Thus, we focus on the mediating role salience plays

Table 3. Example Excerpted From EXCEL Macro That Calculates the Salience Using the ID3.

Entropy Calculation (Based on the Example of the Work Team in Table 1)				
Step 1: Entropy of the Total System				
Entropy About Social Categorization Being In-Group vs. Out-Group				
		Number of Cases	Total Number of Cases	
In-group		12	30	0.4
Out-group		18	30	0.6
	In-group	–0.40	–1.32	0.53
	Out-group	–0.60	–0.73	0.44
Total entropy (S)				0.97
Step 2: Salience of Race Identity				
Data Subset for "Same Race" Branch Testing for Sex				
		Number of Cases	Total Number of Cases	
In-group		10	20	0.5
Out-group		10	20	0.5
Entropy for "same" race branch	In-group	–0.50	–1	0.50
	Out-group	–0.50	–1	0.50
			Total	1.00
Data Subset for "Different Race" Branch Testing for Sex				
		Number of Cases	Total Number of Cases	
In-group		2	10	0.2
Out-group		8	10	0.8
Entropy for "different" race branch	In-group	–0.20	–2.32	0.46
	Out-group	–0.80	–0.32	0.26
			Total	0.72
Total entropy (system, race)			0.91	bits
Information gained by testing race is: Entropy (S) – entropy (S, race)	0.06	bits		
Step 3: Salience of Sex Identity				
Data Subset for "Same Sex" Branch Testing for Race				
		Number of Cases	Total Number of Cases	
In-group		7	14	0.5
Out-group		7	14	0.5
Entropy for "same" sex branch	In-group	–0.50	–1	0.50
	Out-group	–0.50	–1	0.50
			Total	1.00

(continued)

Table 3. (continued)

Data Subset for "Different Sex" Branch Testing for Race				
		Number of Cases	Total Number of Cases	
In-group		5	16	0.31
Out-group		11	16	0.68
Entropy for "different" sex branch	In-group	–0.31	–1.67	0.52
	Out-group	–0.69	–0.54	0.37
			Total	0.90
Total entropy (system, SEX)			0.94	
Information gained by testing sex is: Entropy (S) – entropy (S, sex)	0.03			
Step 4: Salience of Sex × Race Identity				
Data Subset "Same Race" Branch				
Data Subset for "Same Race" and "Same Sex" Branch				
		Number of Cases	Total Number of Cases	
In-group		5	8	0.62
Out-group		3	8	0.37
Entropy for "same sex" branch	In-group	–0.63	–0.67	0.42
	Out-group	–0.38	–1.41	0.53
			Total	0.95
Data Subset for "Same Race" and "Different Sex" Branch				
		Number of Cases	Total Number of Cases	
In-group		5	12	0.41
Out-group		7	12	0.58
Entropy for "different" sex branch	In-group	–0.42	–1.26	0.53
	Out-group	–0.58	–0.77	0.45
			Total	0.98
Total Entropy (System same race, sex)			0.97	
Data Subset "Different Race" Branch				
Data Subset for "Different Race" and "Same Sex" Branch				
		Number of Cases	Total Number of Cases	
In-group		2	6	0.33
Out-group		4	6	0.66
Entropy for "same sex" branch	In-group	–0.33	–1.58	0.53
	Out-group	–0.67	–0.58	0.39
			Total	0.92

(continued)

Table 3. (continued)

Data Subset for "Different Race" and "Different Sex" Branch				
		Number of Cases	Total Number of Cases	
In-group		0	4	0
Out-group		4	4	1
Entropy for "different" sex branch	In-group	0.00	0	0
	Out-group	−1.00	0	0.00
			Total	0.00
Total entropy (system different race, sex)			0.55	
Entropy (S, race × sex)	0.83			
Information gained by testing the interaction race × sex				
Entropy (S) – Entropy (S, race × sex)	0.14			

in the relationship between diversity and team performance (self-reported performance by team members in this case).

Following the same social categorization analysis as illustrated in our prior example, we assess the moderating role of charismatic leadership, as charismatic leadership might invite diversity-blind responses, multicultural responses, or reduce intergroup-biased responses. Thus, we test the following three alternative responses.

Diversity-Blind Response

A first reading of the evidence that charismatic leadership is associated with more positive/less negative effects of diversity is that the effective leadership of which measures of charismatic leadership are presumably indicative (cf. van Knippenberg & Sitkin, 2013) invites diversity-blind responses: when charismatic leadership is low, diversity leads to categorization salience, whereas with high charismatic leadership, diversity does not predict category salience. Effective leadership blinds team members so that they do not "see" these differences because of personal identification with the leader. Diversity-blind responses would thus be evidenced in an interaction effect of diversity/faultlines and charisma predicting categorization salience.

> *Diversity-blind hypothesis:* Diversity is positively associated with categorization salience and this effect is weaker when group members perceive the leader as high, rather than low, on charisma.

Intergroup-Biased Response

Alternatively, charismatic leadership may reduce intergroup-biased responses to salient categorizations. In this case, diversity leads to salient categorizations, and under conditions of low charismatic leadership, salient categorizations invite intergroup bias. Under conditions of high charismatic leadership, however, these intergroup biases are attenuated or eliminated because group members begin to transcend their own short-term self-interest for the sake of their workgroup. Thus, from an intergroup-biased response perspective, we would expect that diversity has a main effect on categorization salience, and that categorization salience and charisma predicting group performance have an interaction effect.

Intergroup-biased hypothesis: Categorization salience will be negatively related to evaluations of team performance and this relationship will be weaker when group members perceive their leader as high, rather than low, on charisma.

Multicultural Response

Alternatively, when differences are perceived they can also be embraced because of their associated informational value and result in positive performance (van Knippenberg et al., 2004). Such a multicultural response perspective would imply that charismatic leadership invites positive responses to salient categorizations, because team members pay attention to the informational value associated with their differences. Like intergroup-biased responses, this implies effects of diversity on categorization salience and of salience-by-charismatic leadership interaction on performance. Importantly, we expect here positive effects of salience, rather than absence of negative effects, on group performance when the leader is high in charisma.

Multicultural hypothesis: Categorization salience will be positively related to evaluations of team performance, and these relationships will be stronger when group members perceive their leader as high, rather than low, on charisma.

Sample and Procedure

The sample for this illustration was drawn from employees working at a large international manufacturing company located in upstate New York (Plant A headquarters) and New Jersey (Plant B). The company was a leader in the development of products and systems that sense, transmit, record, store, and retrieve data, ranging from the wave forms of the human brain or heart to seismic activities. Meetings were organized to inform employees about the general purpose of the study and to ensure that their participation was voluntary and confidential. Following Kozlowski and Bell (2003), we define work teams as groups that "(a) are composed of two or more individuals, (b) who exist to perform organizationally relevant tasks, (c) share one or more common goals, (d) interact socially, (e) exhibit task interdependencies (i.e., workflow, goals, outcomes), (f) maintain and manage boundaries, and (g) are embedded in an organizational context that sets boundaries, and constrains the team, and influences exchanges with other units in the broader entity" (p. 6). We determined work unit memberships from organizational charts, departmental reports (e.g., staffing report from Human Resources) and organizational layouts. We verified this information in interviews with supervisors from each department. We met with each superior to set a group appointment to explain the nature of the study and distribute the survey. Our sample included such work groups as network development, corporate communication management, corporate information systems, industrial sales and marketing, product development, quality control management, and top management.

Participants completed the survey in a meeting room next to the work setting during regular working hours and returned the surveys directly to the lead researcher. Demographic data for each individual participant was recorded from internal archival data. A total of 239 individuals in 38 work teams completed the survey. The demographic description of the sample was as follows: 41% was male, 91% White, 6% Black, 1% Hispanic, and 2% Asian, their average age was 51.7 years, and their average tenure in the company was 10.39 years.

Measures

Salience of social categorizations was measured using the ID3 algorithm as described in the earlier example.

Team diversity. Blau's (1977) index was used to calculate team diversity with respect to sex and race. This index measures the degree to which there are a number of categories in a distribution and the dispersion of the team members within these categories. This index can be expressed as $1 - \sum P_i^2$ where P is the proportion of group members in a category and i is the number of different categories represented on a team. We decided to use the Blau's index to operationalize our measure of sex and race diversity because it is consistent with the categorical scale of our demographic variables. We argue that differences in sex and race can be used as the basis for social categorization but also can be a source of valuable information for the team. This multiple conceptualization is consistent with the CEM (van Knippenberg et al., 2004), which holds that "all dimensions of diversity may have positive as well as negative effects" (van Knippenberg et al., 2004, p. 1008). According to Harrison and Klein's (2007) typology, this rationale implies that we conceptualize sex and race as both separation and variety.

Sex/race faultline. We used Shaw's (2004) index based on a chi-square statistic to measure the two-dimensional faultline of sex and race. In the case of two attributes, this is equivalent to a cross-table correlation coefficient that captures the correlation between positions on two dimensions of diversity. It captures both within-subgroup similarities and between-group differences. Higher correlations indicate stronger faultlines. For example, let us take the case of sex-by-race faultline in two teams of five people. The three female members of Team 1 are Hispanic, the two male members of Team 1 are White, and therefore the alignment between the two variables (faultline) is 1. In contrast, in Team 2, there is variation in race within gender. The 2 female members of Team 2 are Hispanic, but the third is White, the faultline is weaker ($r = .40$).

Perceived task performance. Measures of team task performance were obtained by having all team members complete the Van de Ven and Ferry's (1980) work group performance scale, including seven factors: quantity and quality of performance, number of innovations, reputation for work excellence, goal attainment, efficiency, and morale. All items were evaluated on a 5-point scale, ranging from 1 = *far below average* to 5 = *far above average*. Items were combined into a scale by computing the mean on raw scores. The reliability index was .82. Scores were aggregated at the team level.

Charismatic leadership. Team members evaluated their formal leader using the charisma scale, taken from the Multifactor Leadership Questionnaire (Bass & Avolio, 1990), with a reliability index of .88. A sample item is, "He/she goes beyond his/her self-interest for the good of our group." Scores were aggregated at the team level.

Controls. We included *team size* as a control because it is often argued that team size can impact how team members perform and interact (Curral, Forrester, Dawson, & West, 2001). We also added a *team plant* dummy as a control to our analyses as the work climate within a plant may impact team processes. *Team tenure* or the average amount of time team members have spent with the team has long been considered to have an influence on group development (Pfeffer, 1983; Weick, 1969) and has been shown to affect team outcomes (Finkelstein & Hambrick, 1990; Hirst, 2009). As a result, we believe it warrants inclusion as a control variable in our analyses. Finally, *task interdependence* is the degree to which accomplishing the goal of the team requires completion of interconnected or interdependent subtasks. More interdependent tasks require higher levels of collaboration and information sharing among team members (Bell & Kozlowski, 2002). In teams with less interdependent tasks, there is less of a need for collaboration and information sharing; hence the level of interdependence within a team may impact team outcomes. Task interdependence was measured using a four-item survey developed by Van de Ven, Delbecq, and Koenig (1976). The task interdependence scale indicates what percentage of the total work within their unit flows in each of the ways as shown by each of the figures (see Appendix A): (a) independent work flow, (b) sequential work flow, (c) reciprocal work flow, and (d) team work flow. These different work flow cases were assigned values of 0, 3, 6, and 9, respectively. Then, the scaled value corresponding to the most predominant work flow in each group was used as the measurement of task interdependence. The average was 52.59 and standard deviation was 18.28, suggesting that indeed most groups were moderate to high interdependence but that there was enough variance on this variable among the groups (see Appendix A).

Results

Measurement Model

Both charismatic leadership and team performance were subjective ratings by team members, and therefore we ran a confirmatory factor analysis to test whether a two factor charismatic leadership—team performance solution had a good fit and a better fit than the one-factor alternative model. The two factor model had good fit, $\chi^2(20) = 60.49$, CFI = .96, SRMR = .08, that was superior to that of the one factor model, $\chi^2(20) = 715.14$, CFI = .43, SRMR = .20, according to a χ^2 difference test. All factor loadings were statistically significant and ranged from .42 to .82 for team performance items and .53 to .94 for charismatic leadership items.

Data Aggregation

To assess the suitability of aggregating individual scores of team performance and team leader's charisma to the organizational level, we examined intraclass correlations, that is, ICC(1) and ICC(2), as well as within group agreement, that is, $r_{WG(j)}$ (Le Breton & Senter, 2007). Although these scales already referred to the team level, we checked empirically the appropriateness of aggregating the responses of individual team members to the team level. For team performance, ICC(1) was .05 and ICC(2) was .23; and for team leader's charisma, ICC(1) was .21 and ICC(2) was .62. Team performance had a median $r_{WG(j)} = .85$ and leader charisma had a median $r_{WG(j)} = .75$. Taken together, these results suggested very strong agreement and consistency in responses among group members when evaluating the performance of their team and the charismatic qualities of their leader.

Descriptive Statistics

Table 4 shows correlations, means, and standard deviations for the study variables. The means showed that on average the faultline categorization that captured sex and race simultaneously ($M = .043$) was more salient than the sex categorization ($M = .036$) and the race categorization ($M = .022$). Plant B was more ethnically diverse ($r = .48$, $p < .01$) and on average the salience of the race categorization was also higher in this plant ($r = .30$, $p < .01$).

The zero-order correlations offered preliminary support for the notion that demographic diversity in the team increases the salience of these categorizations. Sex diversity in the team was positively related to the salience of the sex categorization ($r = .38$, $p < .01$). Similarly, race diversity in the team was positively related to the salience of race categorization ($r = .35$, $p < .01$). Furthermore, the sex-by-race faultline was positively related to the salience of the categorization capturing the combination of sex and race ($r = .58$, $p < .01$).

Table 4. Means, Standard Deviations and Correlations.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Team size	7.060	2.850											
2. Team plant B	0.230	0.421	.132										
3. Team tenure	5.720	3.540	.276*	.064									
4. Task interdependence	52.590	18.280	.011	.267*	.035								
5. Sex diversity	0.220	0.194	.256*	–.168	.014	.37**							
6. Race diversity	0.140	0.199	.156	.48**	.005	.189	.000						
7. Sex-race faultline	0.083	0.183	.225	.065	–.077	.162	.456**	.456**					
8. Leader charisma	2.650	0.660	–.046	.151	–.217	.448**	.146	.249	.205				
9. Sex salience	0.036	0.074	–.138	–.211	.072	.203	.379**	–.177	.004	.005			
10. Race salience	0.022	0.047	.103	.299*	–.115	.130	.129	.353**	.129	.315*	–.115		
11. Sex-race salience	0.043	0.083	.279	.049	–.126	.095	.451**	.357*	.151	.582**	.031	.545**	
12. Team performance	3.768	0.360	–.267*	–.009	–.045	.248	–.032	.185	.465**	.087	–.106	–.103	–.173

*$p < .05$. **$p < .01$.

Table 5. Results From Regression Analysis Testing Moderated Mediation Effects of Sex and Race Categorization Salience on Team Performance.

	Model 1	Model 2	Model 3
Variable	Salience Sex	Salience Race	Team Performance
Controls			
Team size	–.34	.14	–.25
Team plant	–.06	.22	–.22
Team tenure	.11	–.08	.17
Task interdependence	.08	.04	.08
Diversity			
Diversity sex	.36*	.19	–.15
Diversity race	–.24	.43*	.10
Moderator			
Charisma	–.10	.14	.61**
Mediators			
Salience sex			–.18
Salience race			–.09
Interactions			
Charisma × salience sex			.28*
Charisma × salience race			–.07
Charisma × diversity sex	.04	–.06	
Charisma × diversity race	.12	–.40	
Adjusted R^2	.03	.03	.36*

*$p < .05$. **$p < .01$.

Regression Analysis: Salience

First, we tested the notion that the salience of social categorizations is influenced by team diversity and diversity faultlines. We conducted hierarchical regression analyses with sex categorization salience regressed on sex diversity and race categorization salience regressed on race diversity, controlling for the size and tenure of the group, plant, task interdependence, and charismatic leadership. The results of the regression analyses are presented in Table 5. Model 1 showed that the coefficient for sex diversity was significant and positive predicting sex categorization salience, $\beta = .36$, $p < .05$. Table 5, Model 2 shows that the coefficient for race diversity was significant and positive predicting race categorization salience, $\beta = .43$, $p < .05$. Table 6, Model 1, shows that there was a positive and significant effect of the sex × race faultline on the salience of the sex × race categorization, $\beta = .61$, $p < .01$.

Next, we tested whether charisma moderated these effects. Model 1, Table 5 shows that this was not the case for the sex diversity-by-charisma interaction or for the race diversity-by-charisma interaction. Given that sex diversity and race diversity had positive relationships with sex categorization salience and race categorization salience respectively, and that these relationships were not moderated by charisma, we can thus rule out diversity-blind responses for sex diversity and race diversity either as a main effect or as an effect moderated by charisma. The question that remained to be addressed below was whether there was a case of multicultural or intergroup-biased responses.

Model 1 of Table 6 shows that the interaction of charisma and sex-by-race faultline predicted salience of the sex × race categorization, $\beta = -.39$, $p < .05$. Figure 2 shows this interaction. The positive relationship between the sex × race faultline and the salience of the sex × race categorization was weaker with higher charismatic leadership, suggesting that charisma inspired diversity-

Table 6. Results From Regression Analysis Testing Moderated Mediation Effects of Sex × Race Categorization Salience on Team Performance.

	Model 1	Model 2
Variable	Salience Sex × Race	Performance
Controls		
Team size	.15	−.20
Team plant	−.08	−.13
Team tenure	−.12	.09
Task interdependence	−.06	.03
Faultline/diversity		
Faultline sex × race	.61**	−.18
Diversity sex	.11	.16
Diversity race	.14	−.10
Moderator		
Charisma	−.05	.56**
Mediator		
Salience sex × race		−.14
Interactions		
Charisma × salience sex × race		.12
Charisma × faultline sex × race	−.39*	
Adjusted R^2	.38**	.28*

*p < .05. **p < . 01.

blindness to some extent. Even so, given that both slopes are positive, these findings also left open the possibility of intergroup-biased or multicultural responses.

Regression Analysis: Team Performance

The notion of multicultural responses would suggest a positive relationship between categorization salience and team performance, whereas the notion of intergroup-biased responses would suggest a negative relationship between salience and performance. Table 5, Model 3, shows that there is no case for a main effect of either sex diversity or race diversity on performance, and Table 6, Model 2, shows that there is no case for a sex × race faultline main effect on team performance. By implication, there is no case to further explore mediation of these main effects.

Moderated intergroup-biased responses would suggest a negative effect of salience on performance that is reduced by charisma. Moderated multicultural responses would suggest a positive effect of salience on performance that is enhanced by charisma (both of course could also obtain; a negative effect with low charisma and a positive effect with high charisma). Accordingly, we tested salience by charisma interactions, as reported in Table 5, Model 3 for the diversity variables and in Table 6, Model 2, for the diversity faultline.

The interaction of sex salience and charismatic leadership predicting team performance was positive and significant, $\beta = .28$, $p < .05$. Figure 3 displays this interaction. The relationship between the salience of the sex categorization and team performance was negative when the leader was perceived as low on charisma, but was positive when the leader was perceived as high on charisma. Thus, with low charisma, responses to sex diversity appeared to be intergroup biased, but with high charisma responses appeared to be multicultural. The interactions with charisma for race categorization salience and sex × race categorization salience were not significant.

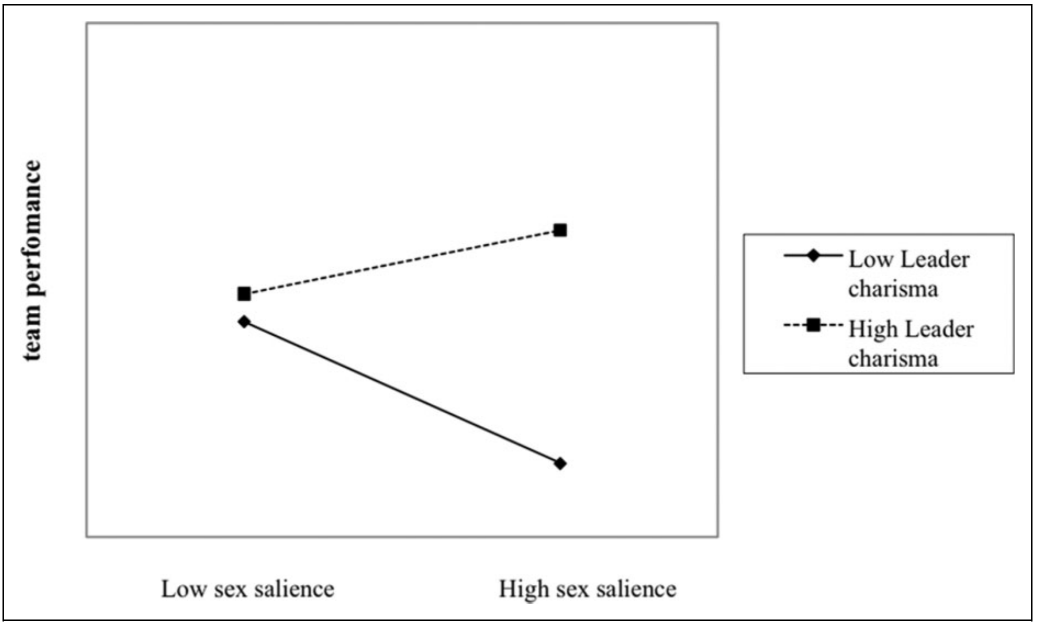


Figure 3. Interaction between sex categorization salience and charisma predicting team performance.

Bootstrapping: The Mediating Role of Salience

To test the moderated mediation model for sex diversity implied by our findings, we used a bootstrap procedure to test the indirect effect of sex diversity on performance through the salience of the sex categorization at –1 SD for charismatic leadership and at +1 SD for charismatic leadership. Mediation is indicated when the indirect effect differs significantly from zero (Shrout & Bolger, 2002). We drew 1,000 random samples with replacement from the full sample and constructed bias-corrected confidence intervals, using the PROCESS macro (Hayes, 2012). This method for evaluating the indirect effect has relatively high statistical power, particularly in the case of small samples (MacKinnon, Lockwood, & Williams, 2004). As shown in Table 7a, for low charisma the indirect effect of sex diversity on team performance mediated by sex salience was –.23, and the 95% confidence interval for this effect excluded zero (–.66, –.01), indicating a significant indirect effect. For high charisma, the indirect effect was .18, and the 95% confidence interval included zero (–.05, .54), indicating a nonsignificant indirect effect. Because the indirect effect was not significant for high-perceived charismatic leadership, we cannot conclude that the response was multicultural. These results supported the conclusion that responses to sex diversity are intergroup biased with low perceived charismatic leadership but not with high perceived charismatic leadership.

As shown in Figure 4, sex × race faultline interacted with charisma to predict salience, and we therefore also tested whether salience subsequently mediated an effect on performance. As shown in Table 7b, for low charisma the indirect effect of sex × race faultline on team performance mediated by sex × race salience was -.22, and the 95% confidence interval included zero (–2.63, 0.33), indicating a nonsignificant indirect effect. Similarly, for high charisma the indirect effect of sex × race faultline on team performance mediated by sex × race salience was .02 and the 95% confidence interval included zero (–1.26, 1.53). Because the indirect effects were not significant,

Table 7a. Conditional Indirect Effects for Sex Diversity × Charisma Interaction.

Level	Charisma	Conditional Indirect Effect	Boot *SE*	Boot LL 95% CI	Boot UL 95% CI
Low = mean − 1 SD	−0.96	−0.22	0.16	−0.66	−0.01
Medium = mean	0.09	−0.02	0.06	−0.16	0.11
High = mean + 1 SD	1.14	0.18	0.16	−0.05	0.54

Note: Number of bootstrap samples for bias corrected bootstrap confidence intervals = 1,000. CI = confidence interval; LL = lower limit; UL = upper limit.

Table 7b. Conditional Indirect Effects for Sex/Race Faultline × Charisma Interaction.

Level	Charisma	Conditional Indirect Effect	Boot *SE*	Boot LL 95% CI	Boot UL 95% CI
Low = mean − 1 SD	−1.02	−0.22	1.25	−2.63	0.33
Medium = mean	0.04	−0.10	0.73	−1.26	0.43
High = mean + 1 SD	1.09	0.02	1.26	−1.26	1.53

Note: Number of bootstrap samples for bias corrected bootstrap confidence intervals = 1,000. CI = confidence interval; LL = lower limit; UL = upper limit.

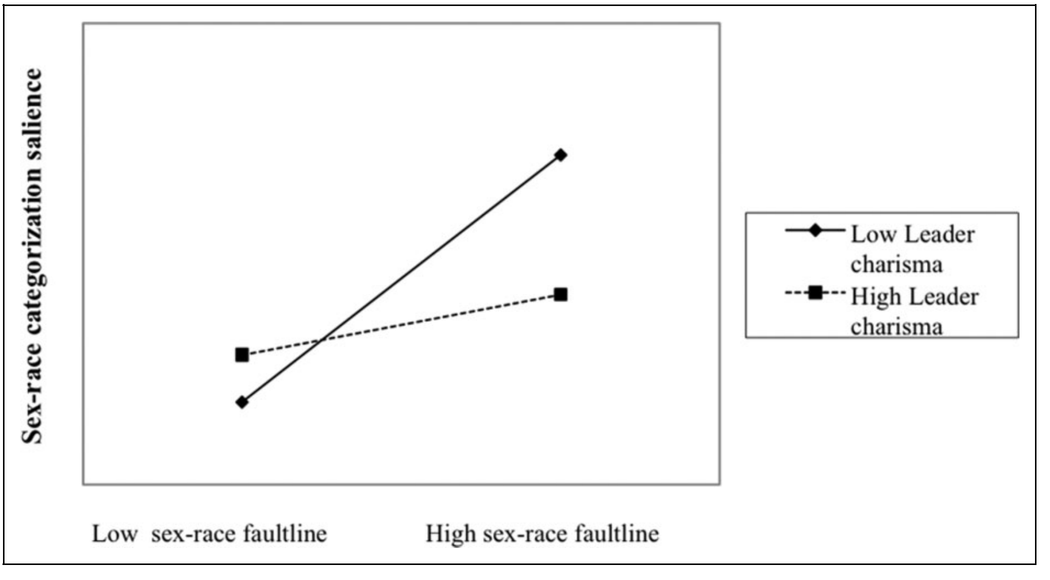


Figure 4. Interaction between sex-by-race faultline and charisma predicting salience of the sex/race faultline categorization.

we can conclude that responses to sex × race faultlines with charisma were limited to diversity blindness with no subsequent effect on performance.

Discussion

Summary of Results

Our findings illustrate the importance of assessing salience to differentiate diversity-blind, multicultural, and intergroup-biased responses to diversity. First, there were no main effects of diversity or diversity faultlines on team performance, but not because diversity or faultlines were not salient. As a main effect, responses were not diversity blind—a conclusion that could not have been reached without explicit assessment of salience. Our results show that demographic sex and race diversity and sex $\times$ race diversity faultline were significantly related to the salience of sex, race, and sex $\times$ race categorization within the team, suggesting that salience of social categories is influenced by objective team diversity and faultlines. Furthermore, we found diversity-blind responses to the sex $\times$ race diversity faultline that was moderated by charisma. Sex $\times$ race categorization salience was lower the more members of the team rated their leaders as charismatic—a pattern of diversity-blind responses as a function of leadership (Figure 1A).

Second, for the diversity faultline, we found evidence for neither intergroup-biased nor multicultural responses (Figures 1B and 1C). Our results show that the indirect effect on team performance mediated by sex-race salience was nonsignificant, both on its own and also when accounting for the moderating influence of charisma on these relationships. This implies that although charismatic leadership preempts intergroup bias by suppressing to some extent social categorization of sex and race (e.g., diversity-blind response), it may not bring about tangible improvements to team performance because it may discourage proactive engagement with diversity as an informational resource.

Third, as moderated effects, we found a buffer to intergroup-biased responses to sex diversity as a function of leadership (but the influence did not take the form of multicultural responses that would inspire performance benefits). With low charismatic leadership ratings, the salience of sex categorization in the team is associated with reduced performance. However, with high charisma ratings, this intergroup-biased response to sex categorization is significantly reduced. Without explicitly assessing salience this effect could not have been distinguished from a diversity-blind response inspired by leadership (i.e., intergroup-biased responses with low charisma, diversity-blind responses with high charisma).

In sum then, this empirical illustration not only shows how the ID3 approach to categorization salience measurement can be applied in diversity research to assess part of the process through which team diversity may affect team performance; it also advances understanding on how doing so can yield important insights that allow more sophisticated testing of theory regarding diversity processes. First, our results urge us to consider categorization salience as a mediator between team diversity and team outcomes. Second, our research is one of the first empirical attempts to explore moderators that may influence the relationship between team diversity and team outcomes. Thus, our moderated mediation approach to diversity is particularly suited to explore the different potential responses to diversity that are at the heart of the CEM proposed by van Knippenberg and his colleagues (2004).

Methodological and Theoretical Contribution of ID3 Salience Measurement

Measuring categorization salience is important to capture one of the most fundamental processes proposed to underlie team diversity effects (cf. van Knippenberg et al., 2004). The ID3 approach to measuring salience we propose here is an important advance over previous self-reports of salience, because it can be applied in concise form (questionnaire length is determined only by number of team members) to multiple dimensions of diversity and diversity faultlines formed by these measures. The inclusion of a salience measure of multiple attributes of diversity might help researchers in four ways. First, although *salience* of categorizations is often evoked theoretically as a key mechanism that mediates demographic diversity and its consequences, measurement techniques of categorization salience are still missing. Harrison et al. (2002), for example, note that "diversity effects rely on perceptions . . . . Yet, these perceptions have rarely been studied in diversity research" (p. 1032). And when studied, perceptual measures are used as markers of the salience of actual diversity (e.g., Harrison et al., 2002), but salience itself is not assessed. As such, a key purpose of our research was a conceptual and methodological presentation on the use of the ID3 algorithm to operationalize categorization salience. This measure can assist in determining the extent to which a set of attributes is psychologically relevant to team members.

Second, our results show that the inclusion of categorization salience measures might help to capture three different responses to diversity among team members in the presence of a moderator—diversity-blind, intergroup-biased, and multicultural perceptions. Our findings provide support for researchers who have argued that moderating factors which influence the relationship between team diversity and team outcomes need to be brought to the fore to gain fine-grained understanding of diversity effects (e.g., Horwitz & Horwitz, 2007; van Knippenberg & Schippers, 2007). The most important implication for future research is that our approach makes it possible to differentiate between the three responses to diversity as proposed by the CEM (van Knippenberg et al., 2004), thus providing a much-needed tool to conceptualize and test more comprehensive theoretical models. By tracing connections between group composition, categorization salience, and group performance, our article helps to gain a more nuanced view of diversity at work.

Third, an additional advantage of the ID3 approach is that it requires interpersonal in- versus out-group or similarity judgments and not self-reports of the use of a social categorization. The latter arguably is more prone to social desirability in responding, because respondents may be more wary of reporting that they rely on demographic characteristics in thinking about their team members. The categorization or similarity judgments are not necessarily free of such concerns—respondents could be aware of the fact that their judgments may reflect, for example, how they perceive different sex or race of others—but because the latter is more implicit than explicit as compared with the existing survey measurement, the concern should at least be weaker.

Finally, our research advances the social categorization and diversity literatures drawing from cognitive process modeling based on machine learning algorithms (e.g., Quinlan, 1986). Scholars showed long ago how computer programs that perform sequences of simple information processing steps could be construed to understand cognitive processes (e.g., Baylor & Simon, 1966). The ID3 represents the structure of the knowledge acquired by the system in the form of a decision tree, similar to the super-ordinate framework of social categorizations. Schrodt (1991) concludes that the ID3 is "the most commonly used machine-learning technique for solving classification problems, it is simple to implement, well understood, robust, and as such has become the artificial intelligence equivalent of stepwise multiple regression as a general-purpose empirical technique" (p. 379). Our analysis suggests that the ID3 machine-learning technique has the potential to provide a much-needed methodological tool to better understand social categorization and its implications for work team diversity. As suggested by Schrodt (1987), "unlike existing statistical techniques, the knowledge representation of structures inductively constructed by [ID3] are plausible models of human inductive theorizing since they fit within the known cognitive constraints of the brain" (p. 379). The ID3 thus does not only have the advantages of being easy to implement and of measurement efficiency, but also of being well-aligned with the cognitive processes it is targeted to measure.

Limitations and Future Research

One limitation of our study is the small sample size of 38 teams. Although this is relatively common in work team research (cf. van Dijk et al., 2012), we have attempted to minimize its impact by reporting adjusted *r*-squares and using two-tailed tests. Also, the use of multisource data avoids inflated correlations due to single method bias. For illustration purposes, we focused on charisma because to date it is the only moderating influence empirically connected to both the effects of diversity and of diversity faultlines. However, we readily acknowledge that conceptually there is little gain in the study of charismatic leadership (van Knippenberg & Sitkin, 2013) and our evidence should be seen as illustrative of the potential of our approach and not as substantively speaking to leadership. Indeed, the key take away here is not about leadership but about the promise of our approach to measuring salience to broadly develop our understanding of moderating influence in diversity's effects for a host of potential moderators.

When introducing new methods in the field, we need to clearly illustrate the advantages of our approach over more traditional alternatives. Thus, we performed additional analyses to demonstrate the value added of the ID3 method over a multilevel approach. We performed multilevel analyses using the continuous ratings of similarity as dependent variable, on a scale from 1 = *very dissimilar* to 10 = *very similar*, and compared the results with the ID3 method. We report these results in Appendix B. Overall, the results of these additional analyses showed negligible effects on the individual level and suggested that similarity ratings operate at Level 2. We found strong between dyadic effects, signaling that there is a strong agreement between the two members in each dyad about how (dis)similar they are with one another. Regarding groups, we found a weak between-group effect. Taken together, these findings suggested that the multilevel model procedure using continuous individuals' ratings of similarity as dependent variables captures the well-established effect of relational demography (similarity in sex and race) on perceived similarity. This is the link between objective diversity and perceived diversity (Harrison et al., 2002). However, the multilevel approach did not capture the effects of the interaction between the demographic attributions in combination. Thus, these results support the value added contribution of the ID3 salience measure over the multilevel approach for three reasons. First, the structural demographic properties of the team (i.e., at the team level) did not relate with individual ratings of similarity at individual level. Second, multilevel models with similarity at the individual level as dependent variable did not capture the effects of the interaction between the demographic attributes in combination. Finally, because modeling between level interactions in multilevel models with dependent dyads is problematic (e.g., Gooty & Yammarino, 2011), we think that it is reasonable to offer alternatives, such as the ID3, that account for categorization salience as a group-level phenomenon.

We also acknowledge that the ID3 salience method may look complicated at first and it may be hard to see the implications for using this new methodology. To appreciate better the broader application of the ID3 salience measurement in diversity research, we have computed the salience for two additional groups in our sample. Let us discuss the example with three groups with moderate (Team 1, which coincides with our earlier example), low (Team 2), and high (Team 3) levels of salience. These examples are presented in Appendix C. Comparisons across groups showed that actual diversity of the group did not translate automatically into categorization salience (compare values of salience vs. diversity with respect to the two social categories and their combination). The three teams had only one member of a minority ethnicity. However, while group members' social categorization in Teams 1 and 3 differed depending on their race, it did not affect them in Team 2. In Team 1, from the twenty social categorization judgments of the "same race" branch 50% were out-groups, whereas within the twelve observations of the "different race" branch 80% were out-groups. This means that to understand social categorization judgments in Team 1, race was important and members were more likely to judge one another as belonging to the out-group when the other dyadic member was different, rather than similar. In contrast, in Team 2, in- versus out-group categorization judgments in both branches, "same race" and "different race" were very similar (7 out of 12 observations were out-groups in the "same race" branch, 58%; and 5 out of 8 observations were out-groups in the "different race" branch, 63%). Consequently, in Team 2 race was not important to understand how people were categorizing their team colleagues. Also, in some groups, information was gained in terms of how team members categorize others by using faultline information (i.e., Teams 1 and 3) whereas in Team 2, the faultline did not have predictive power beyond the information of each social category separately. Finally, it was also important to notice that demographic differences within teams did not translate into salience automatically. A careful look at the values of Team 1, for example, revealed that out of six members, two were female and one Black. However, in this team sex is less salient than race to predict categorization judgments. Taken together, these examples show how future research can make more fine-grained predictions of diversity effects when information provided by the relative salience of diversity attributes within teams and across teams are taken into account.

We limited our analyses to the Big Two of diversity. We illustrated the ID3 application with a focus on sex diversity, race diversity, and the sex $\times$ race faultline, where its advantage over questionnaire measurement (i.e., which would require the repetitive use of the same scale three times) is still relatively modest. With a growing number of diversity attributes, however, its advantage becomes increasingly apparent, because the ID3 application would not require more questionnaire items—only additional information about group members (i.e., functional background, educational background, tenure, age, etc.). The limitation of using the ID3 approach is more in the size of group (i.e., groups need to have at least four members), but because team diversity research usually limits itself to small to modest-sized groups (cf. van Dijk et al., 2012) this should in practice not be an issue.

As with any other methodological approach, the ID3 still requires that researchers make important conceptual decisions when designing their studies. First, researchers need to decide which diversity categories to include in the ID3 algorithm based on theory. Then, the ID3 approach can serve to judge the extent to which the selected set of diversity categories, indeed serve to explain variance of the categorization judgments (in- vs. out-groups) within a team (i.e., assessing the information gained by the system when taking into account a particular attribute). The ID3 cannot preclude the existence of omitted diversity attributes. Just like the study of diversity and diversity faultlines, the ID3 approach relies on attributes identified by the researcher, and thus does not allow for the "emergent" identification of relevant influences. Because the ID3 approach is based on responses as they can be related to diversity attributes identified by the researcher, this also means that when a diversity attribute covaries with another unmodeled variable, influences of that unmodeled variable could be attributed to the diversity attribute captured in the ID3 approach (i.e., the "third variable" problem). In that sense too, the ID3 approach to measuring salience suffers from the same problems as measures of diversity and diversity faultlines in that it does not directly captures *perceptions* of diversity attributes. From that perspective, for single-attribute salience a measure that explicitly addresses perceptions, like Randel's (2002) gender salience measure may be preferable (even when it should be noted that questionnaire measures like Randel's are more likely to contaminate measurement through priming). As soon as more than one attribute comes into play, however, the advantage of measurement parsimony of the ID3 approach should make it preferable to any perceptual rating of salience questionnaire.

Another important decision to be made by researchers considering their specific context is how to categorize their predictor variables. The organizational demography literature has considered certain categorical variables such as race, sex, occupational level, organizational role, educational background, career level, or legal marital status to understand interpersonal processes. However, in some cases researchers need to decide how to partition their variables. For example, organizational role can be categorized as positions with people responsibilities = 1, and without people responsibilities = 0; or with more subcategories (1 = junior positions, 2 = middle-level positions, and 3 = seniors.) Scholars have to decide both which predictors to use and how to categorize them, before being able to perform the ID3. Another limitation refers to how to include continuous predictors in the ID3 algorithm. While decision trees tend to use categorical features (Kotsiantis, 2007), scholars might need to include continuous predictors, such as age or personality traits, to better understand social categorization. Although the ID3 algorithm does not directly deal with continuous predictors, the machine learning literature suggests using entropy-based methods to identify ways to extend the ID3 to do so (Kohavi & Sahami, 1996). There is no universal best approach to do so, and researchers might to consider several aspects before categorizing their continuous variables and using them in the ID3 decision tree. An example of how to apply the ID3 approach with a continuous attribute, age, as well as the methodological and conceptual recommendation on ways to determine partitions for discrete intervals of continuous variables are described in Appendix D.

Finally, the ID3 approach is most suited to surface-level diversity variables as commonly assessed in faultline research, because these attributes are more likely to impact in-group versus out-group categorizations. In practice, we also acknowledge that the referent group matters in social categorizations (see van Knippenberg, van Ginkel, & Homan, 2013). For example, management faculty members in a business school may find age, gender, and race irrelevant to social categorization judgments when using the whole university a referent group (rather than the business school). When applying the ID3 approach, the information gained when considering the demographic attributes might be negligible; hinting to the researchers the need to seek other factors that explain social categorization judgments, such as departmental membership, which can then be incorporated to the ID3 algorithm.

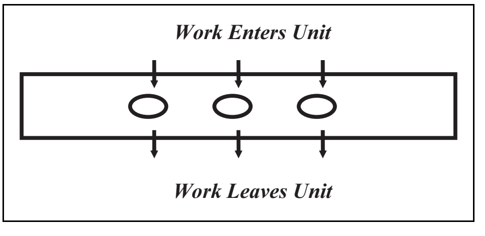
In conclusion, we believe that the ID3 measure of salience should be viewed as an additional resource for understanding group diversity and its consequences at work. It is not a substitute for other measures. To be most useful, the ID3 salience must be integrated within a multidimensional picture of diversity and we would encourage future researchers to do so. Ours is a first step attempt to measure team categorization salience of a combination of attributes, and we hope that it opens avenues to explore diversity effects at work in a more nuanced and comprehensive way.

Appendix A

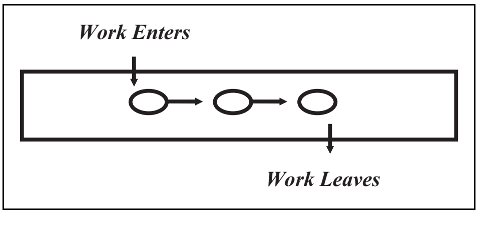
Task Interdependence Scale Rated by the Supervisor

Please indicate what percentage of the total work within your unit flows in each of the ways as shown by the following figures and descriptions. Note that the four answers must add up to 100%.

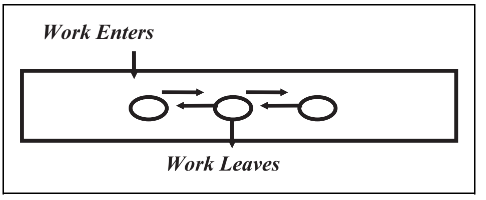
Independent work flow case. Work and activities are performed by your immediate subordinates and do not flow between them: ____________ %



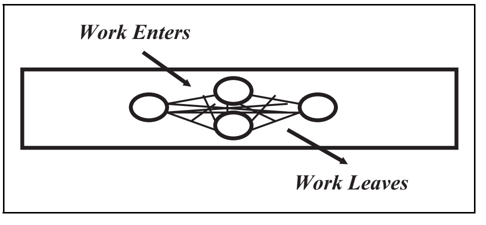
Sequential Work Flow Case. Work and activities flow between your immediate subordinates, but only in one direction: ______________ %



Reciprocal work flow case. Work and activities flow between your immediate subordinates in a reciprocal "back and forth" manner over a period of time: ____________ %



Team work flow case. Work and activities come into your unit and your immediate subordinates diagnose, problem-solve, and collaborate as a group at the SAME TIME to deal with the work: ______________ %



Appendix B

Results from Multilevel Analysis with Continuous Ratings of Similarity as Dependent Variable

We performed additional multilevel analyses using continuous ratings of similarity as dependent variable, on a scale from 1 = *very dissimilar* to 10 = *very similar*). Our data are organized in three levels: individuals nested in dyads nested in groups. However, each individual is not univocally related to one dyad, since each individual is rating every single other member in his/her team. The nonindependence of the observations between dyads requires the use of random coefficient modeling (RCM; Gooty & Yammarino, 2011) and has substantive and methodological implications that we needed to consider. We organized our results in four subsequent steps: (a) preparation of the database to run multilevel analysis with dependent dyads, (b) test of the appropriate level of analysis to examine similarity ratings, (c) test of the multilevel models with individual, dyadic, and group-level predictors, and, finally, (d) exploration of the interactions between our dyadic and group levels.

Step 1. Multilevel data structure. First, we created a database with repeated observations for each participant (i.e., 1,545 observations and 778 dyads, since each participant rated each other team member in terms of similarity). We used cross-classification (Raudenbush & Bryk, 2002) to assess dependent group dyads, such that the nonuniqueness of individual observations in multiple dyads can be accounted by cross-classifying each individual by both the dyad and the group.

Step 2. Testing the appropriate level of analysis. Prior to running the multilevel models that included individual, dyadic, and team variables as predictors of similarity ratings, we needed to assess at what level (i.e., groups, dyads, and individuals) we had significant variance in our dependent variable (i.e., similarity ratings; Hofmann, 1997). To assess this, we first conducted a basic one-way random-effects ANOVA model using the xtmixed command in Stata. We used similarity as the dependent variable and model groups (i.e., Group ID) as Level 2 (Model 1: observations nested in groups). This model has 1,545 observations grouped in 36 teams.[5] An LR test comparing this model to a non-clustered simple linear regression model indicated that the latter should be rejected in favor of Model 1 ($\chi^2 = 23.00$, $df = 1$, $p < .00$). This provides evidence, aligned with our assumption, for significant between-group variation in similarity ratings, although relatively weak, that is, 5% of the variance of our dependent variable existed between work groups.

We next introduced the between-dyad level (Model 2: observations nested in groups, nested in dyads). We performed an LR test to compare Model 2 versus Model 1, and results showed that Model 1 is rejected in favor of Model 2: LR $\chi^2 = 151.19$, $df = 1$, $p < .00$. Aligned with our predictions, results provided evidence for strong between-dyad variation in similarity ratings and indicated that 43% of the variance of our dependent variable existed in between dyads. This supports the notion of salience as a shared phenomenon among team members.

Finally, we introduced the individual level in Model 3 since we have repeated observations (i.e., each individual in our sample is rating all other members in his or her group; Model 3: observations nested in groups, nested in dyads, nested in individuals). Also aligned with our predictions, the results of the LR test comparing Model 3 to Model 2 showed that Model 3 did not significantly improve the explained variance of similarity ratings in comparison to Model 2 ($\chi^2 = 1.59$, $df = 1$, *ns*). These results indicated that the individual factors are overshadowed by the group-, and especially, the dyadic-level factors.

Step 3. Testing multilevel models' fixed effects of individual, dyadic, and group-level predictors. Based on our previous analysis, we run an RCM via cross-classification for dependent dyads. Both Kenny, Kashy, and Cook (2006) and Gooty and Yammarino (2011) warned about the statistical challenges of dependent dyads in organizational research. They highlighted this difference to test predictions in datasets with interdependent data, such as ours. A typical Level 1 and Level 2 model with individuals (i) in dyads (j) cross-classified by groups (k) is as follows:

$$\text{Level 1: } Y_{ijk} = p_{0jk} + e_{ijk} \tag{1}$$

$$\text{Level 2: } p_{0jk} = y_o + b_{00j} + c_{00k} \tag{2}$$

Equation 2 accounts for the dependency between dyads and group simultaneously (i.e., individuals in dyads and groups), as it builds in the factor term $\{c_{00k}\}$ modeled simultaneously with the dyad factor term $\{b_{00j}\}$. We performed multilevel model in this way using the xtmixed command in STATA (Cameron, Gelbach, & Miller, 2011) and specifying the multiway clustering between dyads and groups as follows:

|| _all: R.GroupID || DyadID:, mle

We entered individual-level predictors (sex and race), dyadic predictors (i.e., we dummy coded whether each "target" other was the same or different on sex and race; 0 = *similar* and 1 = *different*) and group-level predictors: group diversity sex and race (using Blau's team diversity indices, 1977) and sex × race faultline (using Shaw's, 2004, index). Results showed that both sex (–.60, $p < .01$) and race (–.91, $p < .01$) dyadic differences were negatively related to similarity ratings, but not their interaction. None of the individual factors was significantly related to similarity ratings. Moreover, none of the group-level variables significantly predicted similarity ratings.

Step 4. Additional analysis: Exploring the interaction between dyadic and group levels. The results above showed that none of the team-level variables were predicting ratings of similarity. The fact that the group-level variables did not have a direct effect on individual ratings of similarity was hardly surprising. In fact, why should they? Since we had similarity ratings at dyadic level, the group average of the individual similarity ratings in a group whose members are similar in sex might not necessarily be higher than the group average in a dissimilar group with respect to sex.

Our ID3 approach however does account for the fact that for different groups, sex has a different importance in the categorization processes of members as pertaining to in- versus out subgroups and thus aligns the diversity variables at the team level with a team-level outcome variable (i.e., ID3 identity salience). This presupposes that sex and race differences might have different effects on similarity ratings depending on the team members are a part of (i.e., how relevant is sex to classify individuals in a particular group as in- vs. out-groups, which is captured by our ID3 salience calculated for each team in our sample?). Using a multilevel approach with an individual-level outcome variable might not be a straightforward way to address this issue empirically. Thus, we performed additional analyses to assess whether certain structural demographic properties at the dyadic level relate to different levels of similarity ratings depending on the group.

For doing that, we divided our data into dyads with no differences (i.e., *sex difference* = 0, *race difference* = 0), with sex differences only (i.e., *sex difference* = 1, *race difference* = 0), with race differences only (i.e., *sex difference* = 0, *race difference* = 1), and with both sex and race differences (i.e., *sex difference* = 1, *race difference* = 1). If our rationale holds, we would have significant between-group variance for each of the four dyadic structural characteristics. We performed one-way ANOVAs to test this. In dyads where there were neither sex nor race differences ($n = 831$), the between group variance was significant ($F = 2.262$, $df = 34$, $p = .001$) and 9% of the total variance of similarity ratings was between groups. In dyads where there were only race differences ($n = 155$), the between group variance was also significant ($F = 4.501$, $df = 16$, $p = .000$) and 34% of the total variance in similarity ratings was between groups. In dyads with only sex differences ($n = 428$), the between group variance was significant ($F = 1.999$, $df = 23$, $p = .004$) and 10% of the total variance in similarity ratings occurred between groups. Finally, in dyads different both in sex and race ($n = 55$), the between group variance was significant ($F = 3.45$, $df = 9$, $p = .003$) and 40% of the variance occurred between groups. These results suggest that the extent to which race dyadic differences relate to similarity ratings is more group context specific than sex differences. Overall, they indicated that the same demographic differences might mean different things for different groups in terms of similarity categorization processes.

Appendix C

Social Categorization Data of Teams 2 and 3

Examples of Group Configurations With Different Values of Sex × Race Salience

Person 1	Person 2	Person 3	Person 4	Person 5	Person 6	Person 7	Entropy (S)	Entropy (S, Race)	**Race Salience: Entropy (S) – Entropy (S, RACE)**	Total Entropy (System, SEX)	**Sex Salience: Entropy (S) – Entropy (S, SEX)**	Total Entropy (S, RACE × SEX)	**Salience Race × Sex: Entropy (S) – Entropy (S, RACE × SEX)**	Diversity Sex (Blau's Index)	Diversity Race (Blau's Index)	Race × Sex Fault-line (Shaw's Index)	
Team 1: Team used in the illustration																	
White Male	White Male	Black Male	White Female	White Female	White Male		0.97	0.91	**0.06**	0.94	**0.03**	0.83	**0.14**	0.44	0.28	0.32	28
Team 2: Team with a lower ID3 sex × race salience																	32
White Male	White Female	White Male	Asian Male	White Male			0.97	0.97	**0.00**	0.95	**0.02**	0.55	**0.02**	0.44	0.49	0.25	
Team 3: Team with a higher ID3 sex × race salience																	
White Male	White Male	White Female	White Male	White Male	Asian Male	White Male	0.92	0.82	**0.09**	0.82	**0.09**	0.55	**0.26**	0.36	0.20	0.17	31.00

Note: In bold, the category salience values calculated using the ID3 algorithm.

Appendix D

Hypothetical Example of the ID3 Algorithm for Age

We illustrate the procedure of the ID3 method to measure age categorization salience. Adapting an example from Thompson and Thompson (1986), we consider a work team of 5 people. Of the total 20 dyads, 15 are in-groups and 5 out-groups (see Table D1). The continuous variable of age used to predict the classification of the 20 cases into in-groups and out-groups can have the following relational values: same (0), younger (–1), and older (1). Scholars might need to consider the following four options when using continuous predictors in the ID3:

1. *Theory-driven cutoff points.* Researchers can decide splitting continuous variables based on theory. For example, generational and adult development theory (e.g., Levinson, 1986) could be used to decide age intervals that capture meaningful inflexion points such as early adulthood or midlife transitions. Consistent with this approach, scholars have in the past measured age as a categorical variable considering meaningful intervals based on existing research (e.g., Aryee, Fields, & Luk, 1999; Finegold, Mohrman, & Spretizer, 2002).
2. *Population norms*. Another way of splitting continuous variables is using publicly available population norms from existing research. For example, in the case of personality, Costa and McCrae (1992) have done an impressive job making available population norms for the Big Five personality traits (e.g., extroversion). Researchers can use, if applicable, these available cutoffs based on norms to convert values of their continuous predictor variables into various discrete bins.
3. *Statistical cutoff points*. Researchers can split continuous variables based on cut-points that split our sample in *i* equal groups in our sample. For instance, individuals can be classified as either high or low in one dimension based on the median score in the sample. In this case, dyads with members both in either the low or high subgroups will be considered similar on the target dimension. We acknowledge that this method might look arbitrary. Alternatively, researchers can create dichotomous variables using the sample median of the dyadic differences on the continuous variables (i.e., say, for example, that the median in our sample is 8 years of difference between the two dyadic members; ranging from 33 to 0). Any dyad within an absolute age difference value of 8 will be considered similar, above the median will be considered older, and below younger. Then, one can regress the dependent variable (i.e., social categorization in in- vs. out-groups in our case) against the newly created dichotomous variable in the total sample. Researchers might then create two additional dichotomous variables with the next closest values above and below the median (i.e., 6 and 10 years of difference, for example) and use them consecutively as predictors of the dependent variable. Then, one can compare the variance of the outcome variable explained by the different dichotomous variables and repeat this procedure until the differences of explained variance of the outcome variable are negligible. Researchers can choose the cutoff that explains the largest variance of the target dependent variable.
4. *Informative cutoff points*. Finally, although the ID3 algorithm does not directly deal with continuous predictors, scholars in the machine learning literature have used entropy-based methods to identify ways to extend this algorithm to do so (Kohavi & Sahami, 1996) and determine partitions for discrete intervals of continuous variables. Indeed, this literature offers various discretization techniques that can be used to identify the most informative cutoff points in particular samples (Chickering, Meek, & Rounthwaite, 2001). For example, let us say that our sample is formed by *i* individuals forming *m* dyads. An attribute in our sample is, Age_i, that has a continuous range of values between 33 and 66 years old. We first

Table D1.

Dyad	Relational Age	Theoretical Cutoff	Cutoff 2	Cutoff 3	Cutoff 5	Cutoff 10 (Median)	Cutoff 14	Cutoff 12	Cutoff 15	Cutoff 17	Social Categorization
1 → 2	10	1	1	1	1	**1**	0	0	0	0	In-group
1 → 3	−2	0	−1	0	0	**0**	0	0	0	0	In-group
1 → 4	12	1	1	1	1	**1**	0	1	0	0	In-group
1 → 5	−5	0	1	1	1	**0**	0	0	0	0	In-group
2 → 1	−10	−1	−1	−1	−1	**−1**	0	0	0	0	In-group
2 → 3	−12	−1	−1	−1	−1	**−1**	0	−1	0	0	In-group
2 → 4	2	0	1	0	0	**0**	0	0	0	0	Out-group
2 → 5	−15	−1	−1	−1	−1	**−1**	−1	−1	−1	0	In-group
3 → 1	2	0	1	0	0	**0**	0	0	0	0	In-group
3 → 2	12	1	1	1	1	**1**	0	1	0	0	In-group
3 → 4	14	1	1	1	1	**1**	1	1	0	0	In-group
3 → 5	−3	0	−1	−1	0	**0**	0	0	0	0	In-group
4 → 1	−12	−1	−1	−1	−1	**−1**	0	−1	0	0	Out-group
4 → 2	−2	0	−1	0	0	**0**	0	0	0	0	Out-group
4 → 3	−14	−1	−1	−1	−1	**−1**	−1	−1	0	0	Out-group
4 → 5	−17	−1	−1	−1	−1	**−1**	−1	−1	−1	−1	Out-group
5 → 1	5	0	1	1	1	**0**	0	0	0	0	In-group
5 → 2	15	1	1	1	1	**1**	1	1	1	0	In-group
5 → 3	3	0	1	1	0	**0**	0	0	0	0	In-group
5 → 4	17	1	1	1	1	**1**	1	1	1	1	In-group

need to examine the potential relational difference age values in our sample at the dyadic level (i.e., a continuous range): Relational Difference Age$_m$ = RDA$_{1,\ min}$, RDA$_2$, RDA$_3$ ... RDA$_{m,\ max}$, where RDA$_{m,\ max}$ corresponds to the maximum relational difference age value for dyad *m* in our sample (i.e., most dissimilar dyad with respect to age; for example, 33 years of difference between the two dyadic members), and RDA$_{1,\ min}$ corresponds to the minimum one (i.e., most similar dyad with respect to age; for example 0 years of difference between the two dyadic members). Then, for each of the RDA$_{1,\ 2,\ \ldots,\ m}$ values, we dichotomize [cutoff_RDA$_{1,\ 2,\ \ldots,\ m}$] as dissimilar dyads [1] with relational differences values up to the RDA$_m$ value, and as similar [0] those that have values smaller than RDA$_m$. For each of these partitions we compute the gain, or gain ratio and choose the partition that maximizes the gain. The gain ratio is an information-based measure that takes into account different cutoff points in relation to the outcome variables. This procedure is described in the literature as a dynamic discretization of continuous predictors, and it is worth noting that the threshold(s) used in one particular group for a particular continuous variable can be different from another one used in a different team. Note that this approach requires a lot of computations. Researchers might opt for exploring splitting points whose relational difference values are closer to the median of the continuous predictor's values, and stop the computations when the information gain starts to decrease.

We use ID3 to determine the gain in information provided by age when deciding whether the social categorization is in-group or out-group. Applying the below entropy equation to the example set in Table D1 gives us a measure of uncertainty (or entropy) about social categorization being in-group or out-group.

Equation 1:

$$\mathrm{H}\,(\mathrm{C}) = -\sum_{\mathrm{i}=1}^{N} \mathrm{p}(ci)\,\log_2 \mathrm{p}(\mathrm{ci})$$

where p(c_i) is the proportion of individuals who belong to each class of the social categorization variable and the log base 2 represents the number of bits needed to represent that many different individuals.

Because there are two possible values for social categorization (in-group and out-group), the probability of having an "in-group" value is 15 out of a total sample set of 20. The probability of an "out-group" value is 5/20. Thus, the entropy of classification for the total data set is

$$\begin{aligned}\mathrm{H}\,(\mathrm{C}) &= -\,\mathrm{p(in-group)}\log \mathrm{p(in-group)} - \mathrm{p(out-group)}\log \mathrm{p(out-group)}\\ &= -\,(15/20)\log_2(15/20) - (5/20)\log_2(5/20) = 0.82 \text{ bits}\end{aligned}$$

This value is the total degree of entropy in the data set. The second step is to determine the amount of information contained in the age category. To calculate the entropy of classification after deciding on age, represented by H(System/age), the ID3 algorithm splits the example set into subsets where each example has the same value of the age attribute.

Although the ID3 algorithm does not directly deal with continuous predictors, the machine learning literature suggests using entropy-based methods to identify ways to extend the ID3 to do so (Kohavi & Sahami, 1996). In this example, we followed this approach and illustrate the information gained by considering the age category at different cutoff points to determine which one is the most informative in our sample. We start with the cutoff closest to the sample relational age median with a cutoff of 10 (in bold in Table 1). We classify absolute relational age values up to 10 as similar, and above 10 as dissimilar (i.e., negative values pertain to the "younger" category and positive to the "older" one). After splitting by age, the "older" branch gives 6 in-group cases, the "same age" branch gives 6 in-group cases and 2 out-group cases, and the "younger" branch gives 3 in-group cases and 3 out-group cases.

Entropy of system before testing = – (15/20) $\log_2$ (15/20) – (5/20) $\log_2$ (5/20) = 0.82 bits

- Entropy for "older" branch = 0
- Entropy for "same" age branch = – (6/8) $\log_2$ (6/8) – (2/8) $\log_2$ (2/8) = 0.82 bits
- Entropy for "younger" branch = – (3/6) $\log_2$ (3/6) – (3/6) $\log_2$ (3/6) = 1 bit

Entropy (S, age) = (6/20) (0) + (8/20) (.82) + (6/20) (1) = 0.63 bits

Information gained by testing age is:

$$\text{Entropy (S)} - \text{Entropy (S, age)} = 0.82 - 0.63 = \underline{0.19 \text{ bits}}$$

Then we performed identical calculations with the rest of potential relational age thresholds in our sample (i.e., 2, 3, 5, 12, 14, 15 and 17; see second column in Table D1) to assess which one is the most informative when predicting social categorization. The information gained by the system with a cutoff of 2 relational age difference was .15, with a cutoff of 3 was .28, with a cutoff of 5 was .24, with a cutoff of 12 was .22, with a cutoff of 14 was .10, with a cutoff of 15 was .07 and with a cutoff of 17 was .03. These results suggested that the most informative relational age threshold in our sample was 3 years difference between the dyadic members. In this example we left aside Options 1 (theoretical cutoff points), 2 (population norms), and 3 (statistical cutoff points); however, it is important to note that once the researcher has decided on the cutoff points of the continuous(s) predictor(s), the ID3 computations required to measure categorization salience are identical regardless of the strategy followed.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study has been partially funded by Research Grant ECO2012-33081 from the Ministry of Economy and Competitiveness. A short version of this article was published in the Best Paper Proceedings of the OB Division of the Academy of Management, 2011.

Notes

1. We thank an anonymous reviewer for suggesting this point.
2. The dichotomization of variables may reduce their observed correlations with other related constructs and thus underestimate their relationships (Law, Wong, & Mobley, 1998). To minimize this problem, we evaluated the distribution of our similarity variable. The fact that our distribution was bimodal and the midpoint of the judgmental continuum was exactly in the midpoint of the scale (M = 4.97, SD = 2.67) suggested that the dichotomization would not challenge the validity of the conclusions presented in our example.
3. That the $\log_2$ (8) = 3 means that 3 bits of information are needed to represent 8 objects. That is, these 8 objects can be coded in binary as 000, 001, 010, 011, 100, 101, 110, and 111 (Garson, 1991).
4. The macro is available online http://margaritamayo.com/files/2016/04/MACRO-ENTROPY-CALCULATION-MAYO-ET-AL.-2016.xlsx. The relative salience of each diversity attribute is calculated automatically. The user needs to follow the instructions and provide the necessary data in the "introducing your data" spreadsheet.

5. Please kindly note that the complete sample for running the multilevel analyses includes 36 teams, as compared to the 38 teams reported in the methods section. This is because 2 groups in our sample had missing data at the individual and/or dyadic levels.

References

Aryee, S., Fields, D., & Luk, V. (1999). A cross-cultural test of a model of the work-family interface. *Journal of Management*, *25*(4), 491-511.

Bass, B. M., & Avolio, B. J. (1990). *Transformational leadership development: Manual for the Multifactor Leadership Questionnaire*. Palo Alto, CA: Consulting Psychologists Press.

Baylor, G. W., & Simon, H. A. (1966). A chess mating combinations program. In *Proceedings of the Spring Joint Computer Conference* (pp. 431-447). New York, NY: ACM.

Bell, B. S., & Kozlowski, S. W. J. (2002). Goal orientation and ability: Interactive effects on self-efficacy, performance, and knowledge. *Journal of Applied Psychology*, *87*, 497-505.

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, *3*, 27-52.

Blau, P. M. (1977). *Inequality and heterogeneity*. New York, NY: Free Press.

Cameron, A. C., Gelbach, J. G., & Miller, G. L. (2011). Robust inference with multi-way clustering. *Journal of Business & Economics Statistics*, *29*(2), 238-249.

Chickering, D., Meek, C., & Rounthwaite, R. (2001). Efficient determination of dynamic split points in a decision tree. *In the Proceedings of the 2001 IEEE International Conference on Data Mining*, San Jose, CA, pp. 91-98.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Crocker, J., Fiske, S. T., & Taylor, S. E. (1984). Schematic bases of belief change. In J. R. Eiser (Ed.), *Attitudinal judgment* (pp. 197-126). New York, NY: Springer.

Curral, L. A., Forrester, R. H., Dawson, J. F., & West, M. A. (2001). It's what you do and the way you do it: Team task, team size, and innovation-related group processes. *European Journal of Work and Organizational Psychology*, *10*, 187-204.

Earley, P. C., & Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal*, *43*, 26-49.

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

Finegold, D., Mohrman, S., & Spreitzer, G. M. (2002). Age effects on the predictors of technical workers commitment and willingness to turnover. *Journal of Organizational Behavior*, *23*(5), 655-674.

Finkelstein, S., & Hambrick, D. C. (1990). Top management team tenure and organizational outcomes: The moderating role of managerial discretion. *Administrative Science Quarterly*, *35*, 484-503.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York, NY: McGraw-Hill.

Garson, D. (1991). A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review*, *9*, 399-534.

Gooty, J., & Yammarino, F. J. (2011). Dyads in organizational research: Conceptual issues and multilevel analyses. *Organizational Research Methods*, *14*, 456-483.

Greer, L. L., Homan, A. C., De Hoogh, A. H. B., & Den Hartog, D. N. (2012). Tainted visions: The effect of visionary leader behaviors and leader categorization tendencies on the financial performance of ethnically diverse teams. *Journal of Applied Psychology*, *97*, 203-213.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety or disparity in organizations. *Academy of Management Journal*, *32*, 1199-1229.

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.

Hayes, A. F. (2012). *PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling*. Retrieved from http://www.afhayes.com/public/process2012.pdf

Hirst, G. (2009). Effects of membership change on open discussion and team performance: The moderating role of team tenure. *European Journal of Work and Organizational Psychology*, *18*, 231-249.

Hofmann, D. A. (1997). An overview of the logic and rationale for hierarchical linear models. *Journal of Management*, *23*, 723-744.

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, *33*, 987-1015.

Jackson, S. E., Joshi, A., & Erhardt, N. L. (2003). Recent research on team and organizational diversity: SWOT analysis and implications. *Journal of Management*, *29*, 801-830.

Joshi, A., & Roh, H. (2010). 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.

Kearney, E., Gebert, D., & Voelpel, S. C. (2009). When and how diversity benefits teams: The importance of team members' need for cognition. *Academy of Management Journal*, *52*, 581-598.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford.

Kohavi, R., & Sahami, M. (1996). Error-based and entropy-based discretization of continuous features. In *KDD-96: Proceedings of the 2nd international conference on knowledge discovery and data mining* (pp. 114-119). Menlo Park, CA: AAAI.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification technique. *Informatica*, *31*, 241-287.

Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. *12*, pp. 333-375). New York, NY: Wiley-Blackwell.

Kunze, F., & Bruch, H. (2010). Age-based faultlines and perceived productive energy: The moderation of transformational leadership. *Small Group Research*, *41*, 593-620.

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

Law, K. S., Wong, C. S., & Mobley, W. M. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review*, *23*, 741-755.

Le Breton, J. M., & Senter, J. L. (2007). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*, 815-852.

Levinson, D. J. (1986). A conception of adult development. *American Psychologist*, *41*(1), 3-13.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99-128.

Nederveen Pieterse, A. N., van Knippenberg, D., & van Dierendonck, D. (2013). Cultural diversity and team performance: The role of team member goal orientation. *Academy of Management Journal*, *56*, 782-804.

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

Pfeffer, J. (1983). Organizational demography. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. *5*, pp. 299-357). Greenwich, CT: JAI.

Pierce, J. R. (1980). *An introduction to information theory*. New York, NY: Dover.

Plaut, V. C., Thomas, K. M., & Goren, M. J. (2009). Is multiculturalism or color blindness better for minorities? *Psychological Science*, *20*, 444-446.

Quinlan, J. R. (1983). Learning efficient classification procedures and their application to chess end games. In R. S. Michalski, J. G. Carbonell, & R. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 463-482). Palo Alto, CA: Tioga.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81-106.

Randel, A. E. (2002). Identity salience: A moderator of the relationship between group gender composition and work group conflict. *Journal of Organizational Behavior*, *23*, 749-766.

Randel, A. E., & Jaussi, K. S. (2003). Functional background identity, diversity, and individual performance in cross-functional teams. *Academy of Management Journal*, *46*, 763-774.

Randenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear model: Applications and data analysis methods*. California: Sage Publications.

Rosch, R. (1975). Universals and cultural specifics in human categorization. In R. Brislin & S. Bochner (Eds.), *Cross-cultural perspectives in learning* (pp. 177-207). Greenwich, CT: JAI.

Schrodt, P. A. (1987). *Microcomputer methods for social scientists*. Newbury Park, CA: Sage.

Schrodt, P. A. (1991). Prediction of interstate conflict outcomes using a neural network. *Social Science Computer Review*, *9*, 359-380.

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

Shin, S. J., Kim, T.-Y., Lee, J.-Y., & Bian, L. (2012). Cognitive team diversity and individual team member creativity: A cross-level interaction. *Academy of Management Journal*, *55*, 197-212.

Shin, S. J., & Zhou, J. (2007). When is educational specialization heterogeneity related to creativity in research and development teams? Transformational leadership as a moderator. *Journal of Applied Psychology*, *92*, 1709-1721.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422-445.

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*, 207-218.

Thatcher, S. M. B., 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.

Thompson, B., & Thompson, W. (1986). Finding rules in data. *BYTE*, *11*(12), 149-158.

Turner, J. C. (1987). A self-categorization theory. In J. C. Turner, M. A. Hogg, P. J. Oakes, & M. S. Wetherell (Eds.), *Rediscovering the social group: A self-categorization theory* (pp. 42-67). Oxford, UK: Basil Blackwell.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Blackwell.

Van de Ven, A. H., Delbecq, A. L., & Koenig, R. Jr. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, *41*, 322-338.

Van de Ven, A. H., & Ferry, D. L. (1980). *Measuring and assessing organizations*. New York, NY: John Wiley.

van Dijk, H., van Engen, M. L., & van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, *119*, 38-53.

van Ginkel, W. P., & van Knippenberg, D. (2008). Group information elaboration and group decision making: The role of shared task representations. *Organizational Behavior and Human Decision Processes*, *105*, 82-97.

van Knippenberg, D., Dawson, J. F., West, M. A., & Homan, A. C. (2011). Diversity faultlines, shared objectives, and top management team performance. *Human Relations*, *64*, 307-336.

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.

van Knippenberg, D., Homan, A. C., & van Ginkel, W. P. (2013). Diversity cognition and climates. In Q. Roberson (Ed.), *The Oxford handbook of diversity and work* (pp. 220-238). New York, NY: Oxford University Press.

van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, *58*, 515-541.

van Knippenberg, D., & Sitkin, S. B. (2013). A critical assessment of charismatic-transformational leadership research: Back to the drawing board? *Academy of Management Annals*, *7*, 1-60.

van Knippenberg, D., van Ginkel, W. P., & Homan, A. C. (2013). Diversity mindsets and the performance of diverse teams. *Organizational Behavior and Human Decision Processes*, *121*, 183-193.

Weick, K. E. (1969). *The social psychology of organizing*. Reading, MA: Addison-Wesley.

Williams, K, & O'Reilly, C. (1998). Demography and diversity in organizations: A review of 40 years of research. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. *20*, pp. 77-140). Greenwich, CT: JAI.

Author Biographies

Margarita Mayo is professor of organizational behavior at the IE Business School in Madrid. She was a Fulbright scholar at Harvard University and professor at the Ivey School of Business. She has published her work in journals such as the *Journal of Applied Psychology*, *Academy of Management Journal*, *Academy of Management & Education*, and the *Leadership Quarterly*. Her research interests revolve around leadership, social identity, authenticity, team diversity, and work-life balance.

Daan van Knippenberg is professor of organizational behavior at the Rotterdam School of Management, Erasmus University Rotterdam, Netherlands. His research interests include leadership, diversity, team performance, creativity and innovation, and social identity. He is incoming editor-in-chief of the *Academy of Management Annals*, editor-in-chief of *Organizational Psychology Review*, and an associate editor of the *Academy of Management Journal*.

Laura Guillén is an assistant professor of organizational behavior at the European School of Management and Technology (ESMT, Berlin), where she teaches in the MIM, MBA, and EMBA programs. Her research interests revolve around issues of leadership, identity, motivation, gender, and diversity. Her work has been presented at numerous international research conferences and published in journals including *Organizational Behavior and Human Decision Processes* and the *Leadership Quarterly*.

Shainaz Firfiray is an assistant professor in the organization and human resource management group at Warwick Business School, University of Warwick, UK. She received her PhD in management at IE Business School, Spain. Her research interests include work-life balance, social identity, diversity, and family firms. Her research has been published in *Human Resource Management* and *Oxford Handbook of Creativity, Innovation, and Entrepreneurship*.