

Status-Based Stereotyping Effects on Military Team Performance

Evidence from an International Military Skills Competition

By Lissa V. Young & Elizabeth L. Wetzler*

Work teams are often touted as a way to increase or enhance organizational performance¹ and yet, both research² and anecdotal evidence suggest that they often fail to do so. While there are several reasons why teams may not succeed in producing optimal performance, one particularly salient issue is the underutilization of team members' knowledge and skills.³ Indeed, it is well documented that team members are often unable to fully capitalize on the team's collective know-how due to members' inability to see beyond their own demographic, be it gender, race or social status, to leverage one another's task-relevant knowledge, skills, and abilities.⁴ Unfortunately, this oversight can leave much of the team's talent untapped, potentially hindering the team's performance.⁵ The cost of this oversight can be significantly deleterious to military operations.

Though this tendency of team members to focus more on colleagues' demographics than their potential talented contributions is common, researchers tend to promote the virtues of team diversity while overlooking the ways that these cognitive classifications (i.e., "stereotypes") can undermine the very benefits that diverse teams may offer. One exception comes from status characteristics theory, in which Berger *et al.* (1972) offer an account of how biased evaluations can occur in team settings. Their theory explores how group members use status differences to drive their expectations for the performance abilities of fellow team members. These performance expectations, in turn, affect the power and prestige orders of these groups,⁶ and these critical group dynamics impact the team's performance. Hence, status characteristics theory offers a compelling lens through which to examine the impact of diversity on teams.

A criticism that has been lobbied against research on stereotyping and its impact on teams is the inconsistent way that team performance has been operationalized and measured.⁷ Team performance studies conducted in natural settings often utilize indirect measures of team performance such as customer satisfaction surveys or supervisor and

* *The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.*

¹ Addison & Haig, 2012 ; Pentland, 2012 ; Devaraj & Jiang, 2019.

² Thatcher & Patel, 2011 ; van Dijk *et al.*, 2012 ; Malik *et al.*, 2019.

³ Hackman, 2002 ; Bachrach *et al.*, 2019.

⁴ Foschi *et al.*, 1985 ; Randolph, 2019.

⁵ Kochan *et al.*, 2003 ; Austin, 2003 ; Hochberg, 2020.

⁶ Berger *et al.*, 1992 ; Foschi, 1992a, 2000 ; Berger *et al.*, 1998.

⁷ Austin *et al.*, 1992.

team members' self-reported assessments.⁸ In the current study, we address this issue by measuring team performance using arguably more objective indices such as time and speed. Our motivation is to reduce the amount of error introduced by indirect or subjective measures of team performance.

Given these various debates, we suggest that alternative methodological approaches may provide opportunities to address previously underexplored aspects and impacts of stereotyping in team settings. For example, the shifting standards model,⁹ if used in conjunction with multilevel structural equation modelling (ML-SEM), may offer us a powerful way to detect and measure multiple levels of stereotyping in teams. As already stated, when encountering lower status group members, people tend to activate their stereotypes. When that happens, expectations for low status group members' competence and performance abilities are implicitly activated. These expectations, in turn, activate judgment standards against which others are evaluated.¹⁰ We propose using the shifting standards model in conjunction with ML-SEM, and more objective measures of team performance, to gain richer insights into the relationship between the multiple status characteristics-based interpersonal biases that occur in team life and that impact team performance.

For this study, we chose an environment where teams are a formalized construct, and the use of teams is highly routinized. We selected a military skills competition, where teams competed against one another in a rigorous three-day event. The selected teams raced against each other, and the clock, in a formal and professional setting. Military teams in the United States are structured so that everyone is assigned distinct roles and responsibilities that are clearly delineated by rank, skills, and abilities. This well-marked interpersonal terrain allows for the study of status-based interpersonal biases in ways that less clearly structured teams cannot.

This project extended previous research in two specific ways. First, it leveraged the strength of structural equation modelling to extend the use of the shifting standards model from the individual to the team level. This added to the theoretical power of the shifting standards model because the model has never been used to examine the impact of team-level shifting standards on team performance. Second, this research measured the impact of stereotyping on team performance utilizing standards of performance that were relatively more ordinal and objective than found in the existing research on teams.

As the workplace and work teams become increasingly diverse, especially within the United States, there has been a tension between the promise and reality of diversity in team performance.¹¹ The pessimistic view, held by theoretical perspectives such as social

⁸ Watson *et al.*, 1993 ; Milliken & Martins, 1996 ; Wells & Aicher, 2013.

⁹ Biernat *et al.*, 1991 ; Biernat & Manis, 1994.

¹⁰ Biernat, Vescio & Manis, 1998 ; Weeks, 2019.

¹¹ Mannix & Neale, 2005 ; Wells & Aicher, 2013.

categorization (Tajfel, 1978), is that diversity creates social divisions (“categories”) based on different and quite visible status characteristics such as gender, race, and rank, and these categories, in turn, activate different expectations for team members’ abilities.¹² Rather than fully utilizing the breadth of perspectives and expertise a diverse team might provide, research has shown that team members tend to rely on stereotypes to interact, thus limiting both the range and legitimacy of others’ contributions.¹³ The more optimistic viewpoint, stemming from literature on information processing in teams, holds that diversity will lead to an increase in the perspectives brought to bear on team tasks, which will result in higher levels of creativity and team performance.¹⁴ Despite the lack of unanimity, researchers do agree that diversity matters, whether it results in positive, negative, or mixed impact on teams and their performance.

Thus, it is arguable that stereotyping can hinder team performance, but a vast majority of the research on stereotyping is yet incomplete because it consistently treats as independent each of the categories of stereotyping. It stands to reason that researchers interested in racism, for example, would conduct experiments focused on race, holding constant other aspects of the subjects’ diversity. This is true also of research on other forms of identity-based stereotyping, such as sexism and ageism. However, by focusing entirely on a single aspect of identity, researchers tend to ignore the inherently multifaceted nature of social identity and lose an opportunity to answer fundamental questions of how the myriad social identities “*function in relation to one another*”.¹⁵

We selected two specific bodies of research found in the stereotyping literature. The first is status characteristics theory,¹⁶ which provides the foundation for this study’s point of view on interpersonal stereotype-based judgment standards in teams. The second body of research is that of the shifting standards model,¹⁷ which provides insight into ways of detecting and measuring the presence of stereotype-based judgments in teams. Each of these bodies of literature will be reviewed as it relates to stereotyping and its impact on team performance.

Status Characteristics Theory

Status characteristics theory¹⁸ addresses the ways that observers ascribe domain-relevant competencies, and therefore expectations for performance, to targets based on abilities associated with their demographic categories. Observers then interact with target individuals according to their biased expectations. A status characteristic is any valued attribute that can imply task competence. An example of a status characteristic is having

¹² Foschi, 1992b ; Newheiser & Dovidio, 2012 ; Hoffman & Musch, 2019.

¹³ Ancona & Caldwell, 1992 ; Mannix & Neale, 2005.

¹⁴ Hoffman & Maier, 1961 ; Polzer *et al.*, 2011 ; Mo *et al.*, 2019.

¹⁵ Bodenhausen, 2010 ; Collins & Bilge, 2020.

¹⁶ Berger *et al.*, 1972.

¹⁷ Biernat & Manis, 1994.

¹⁸ Ridgeway & Berger, 1986 ; Berger *et al.*, 1992.

high or low mechanical ability, being female or male, or being White or Black. There are two classes of status characteristics: specific and diffuse. Specific status characteristics (such as mechanical ability) have sharply defined performance expectations in readily identifiable domains. Diffuse status characteristics (such as race, gender, and class) tend to have limited and general performance expectations and are viewed as relevant to a large, indeterminate number of domains.¹⁹ The complication associated with the use of status-based performance expectations is that, often, the status category is not related to the performance domain. For example, the status category “female” is not a relevant predictor in the performance domain “leader,” and yet females are often judged to be less effective leaders than males,²⁰ and this may have profound implications for diverse teams.

Gender determines expectations for social and professional behaviours of men and women, and these expectations can shape interpersonal dynamics on work teams and can have negative effects on team performance.²¹ Often researchers have found that females are perceived less favourably than males with respect to their appropriateness for leadership and their performance as leaders, and this can have deleterious effects on group performance.²²

Interestingly, researchers suggest that status-based judgments on teams are generated to sustain the power and prestige order that has been established in the group.²³ Foschi *et al.* (1994) found that male evaluators tended to underestimate the competence of women in relation to men, even when their performances were the same or better. In their study conducted at the University of British Columbia, male subjects making hiring choices of female and male job applicants consistently hired males at a higher rate than females, regardless of their performance vitae. Foschi *et al.* (1994) assert that these biased evaluations act to sustain the general prestige order of men and women. Other researchers found similar patterns of interpersonal assessments of competence in mixed gender teams of police officers (Gerber, 2011). Gerber found that women were far more likely to be characterized as possessing fewer effective dispositions for police work than their male counterparts, and this carried over into their ability assessments.

These observations reinforce the idea that once status-based perceptions of competence are established in a team, they will be protected and perpetuated.²⁴ One strategy for doing this involves assessing targets by different standards. Success of low-status group members is assessed through a stricter standard than success of high-status group members.²⁵ The presence of status-based expectations has implications for how team members leverage each other’s capabilities in task-oriented work, and this in turn impacts

¹⁹ Berger *et al.*, 1972 ; Smith *et al.*, 2019.

²⁰ Foschi, 1992a; Eagly & Karau, 2002 ; Roberts & Brown, 2019.

²¹ Banaji & Hardin, 1996.

²² Eagly & Karau, 2002 ; Antino *et al.*, 2019.

²³ Foschi *et al.*, 1985, 1994.

²⁴ Ridgeway & Berger, 1986.

²⁵ Foschi *et al.*, 1994 ; Yzerbyt & Demoulin, 2010.

the team's performance.²⁶ For example, Berger *et al.* (1992) conducted a study where participants were assigned notional teammates of differing status characteristics (i.e., gender, race, education, and occupation). Some participants were assigned high-status teammates (male, White, college graduates in a profession) and others were assigned low status teammates (female, non-White, high school graduates, and minimum wage workers). Because the participants' teammates were notional only, their ability levels could only be inferred by their status characteristics. The participants were then challenged with completing and/or solving a variety of different tasks and puzzles. The participant could solve some of these tasks alone, but many required the help of another person ("teammate"). For each problem or task posed, the researcher asked the participant if they wanted to solve the problem alone or with a teammate. In the form of points, the participant could earn rewards for successfully completing a task or solving a problem, or penalties for failing to do so. The object of the exercise was to accumulate the highest number of points. When asked whether they wanted the assistance of the notional teammates, researchers found that participants actively avoided engaging in team-related tasks with those teammates who were lower status, but often chose to engage the assistance of teammates who were higher status. Researchers assert that participants avoided lower status teammates, ostensibly to avoid performing poorly. This status-based avoidance behaviour potentially has significant implications for team performance because it suggests that the talent of lower status team members might be overlooked in favour of higher-status team members.

Similar research in status-based social categorization has delineated the ways that in-group favouritism leads to discrimination of out-group members, and its subsequent influences on team performance.²⁷ Jehn *et al.* (1999) found that social category diversity detracted from team performance because it increased relational conflict between the in-group and out-group members (see also Zhang & Liu, 2019). In her research she found that unmanaged relational conflict stemming from the establishment of in-groups and out-groups interfered with the ability of the teams to remain focused on tasks. These findings are consistent with much of the literature on diversity and team performance.²⁸

There are many status-based influences on team performance, and there is an exhaustive effort to refine and measure how they impact team performance. What we have learned from this research is that when status-based performance expectations are activated in teams, they can have a negative effect on team performance.²⁹ However, despite the extensive research detailing how specific status-characteristics impact team performance, there is still a gap with respect to how all these different status-based biases might act in

²⁶ Foschi, 1992b ; Hackman & Katz, 2010 ; Roberge & van Dick, 2010.

²⁷ Tajfel & Turner, 1979, 1986 ; Jehn *et al.*, 1999 ; Zanutto, Jehn & Bezrukova, 2011.

²⁸ Roberge & van Dick, 2010 ; Antino *et al.*, 2019.

²⁹ Williams & O'Reilly, 1998 ; Brewer & Pierce 2005 ; Kalkhoff *et al.*, 2020.

concert to influence performance. In the last two decades, we propose the use of the shifting standards model³⁰ in extending this research.

The shifting standards model is a category-agnostic method of detecting and measuring the presence of stereotypes, and if used in conjunction with multilevel structural equation modelling, may provide a richer view of the impact of all sources of stereotyping on team performance. In the section that follows, we describe in greater detail the shifting standards model and its relationship to the status-based judgments described in the literature review already outlined.

The Shifting Standards Model

Biernat & Manis (1994) found that when stereotyped individuals are encountered in stereotype-relevant domains, people's stereotypes are activated. This implicitly activates expectations for the stereotyped group members' competence and performance abilities. These expectations are different than for group members that aren't held to a stereotype. The two different expectations, in turn, activate different judgment standards, against which people evaluate the stereotyped and the non-stereotyped group members. These different assessment standards are activated depending on the target being evaluated, and the task being performed. When evaluating two group members, one who is stereotyped and one who is not, regarding a stereotype-relevant characteristic or ability, people tend to evaluate them with standards anchored in their stereotype-based expectations for each. These different standards are determined by bifurcated expectations of the two group members on the dimension being evaluated.

For example, when a man and woman are being evaluated for their athleticism, different standards of judgments are used to determine how athletic each is. These standards are determined by stereotypes about the disparity between male and female athleticism. If the female is thought to be a good athlete, her abilities are being judged in a different way than the man's. The female's standard is compared to the standard held for female athleticism. When evaluating the man's athleticism, the standard used is that of male athleticism. Consequently, if determined that both the man and the woman are good athletes, it's very likely that completely different mental representations are used for what that means for each of them. It is this disparity in the mental representations of people's judgments that the shifting standards model can detect and measure.³¹

The shifting standards model stems from the idea that stereotyped individuals are judged in stereotype-relevant domains using within-category judgment standards. These standards have a scale with a mean and a range that is expected of the target being judged. In stereotype-laden judgment domains, those means and ranges will shift in accordance with expectations of the target being judged.³² For example, athletics is a domain that

³⁰ Biernat & Manis, 1994 ; Biernat *et al.*, 2011.

³¹ Biernat *et al.*, 1991; Biernat & Vescio, 1993 ; Biernat & Manis, 1994.

³² Biernat & Manis, 1994 ; Biernat, Vescio & Manis, 1998.

typically elicits gender stereotypes. If a man and a woman both run a mile in six minutes, despite the equivalent objective performance, the two runners will be subjectively judged relative to different (male and female) standards, resulting in the man being judged as *average* and the woman being judged as *above average* runners. This shift in the standard against which targets are judged in stereotype-relevant domains is called “shifting standards”.³³

This model is particularly worthy as a way of addressing the shortcomings in current stereotyping research in teams. It has the potential to enable researchers to simultaneously measure a variety of sources of stereotyping on teams. In addition, it is especially useful in this study’s context because it can be combined with military-based assessment measures to detect the presence of status-based biases. For example, in a particularly striking demonstration of shifting standards, researchers observed a random selection of US Army officers assigned to teams of 12 to 15 while attending a professional development course.³⁴ The relevant stereotype present in this domain is gender, because women often are discriminated against with respect to leadership in military settings.³⁵ In this study, officers were asked to judge each of their teammates’ leadership competence utilizing two different types of scales. The first was a more subjective Likert-type scale on which they judged team members in several leadership competencies, and the second was a more objective ordinal ranking scale, on which they rank-ordered their teammates with respect to the same leadership competencies. In accordance with the shifting standards model, on the Likert-type “subjective” scale, men rated women much closer to men on all leadership dimensions, and yet on the ordinal-type ranking scale, men ranked women much lower than men on all leadership dimensions.

For example, in the competency of “decision-making”, women were consistently rated on a Likert-type scale as “average” to “above average” with respect to their abilities as decision-makers. However, when ranked in that same competency against their teammates, they were consistently ranked in the bottom one-third of their teams.³⁶ This systematic *shift* between judgments on relatively more subjective Likert-type scales and relatively more objective ranking scales is evidence of the shifting standards effect, and an indicator of the presence of status-based stereotyping. In this way, the shifting standards model can be used to detect otherwise well-masked stereotyping and prejudice in teams.

Although shifting standards has been used to measure stereotyping on a variety of dimensions by individuals among individual targets, shifting standards has not yet been utilized to examine the relationship between stereotyping and group performance.³⁷

³³ Biernat & Manis, 1994 ; Biernat & Kobrynowicz, 1999 ; Glock & Kleen, 2020.

³⁴ Biernat, Crandall *et al.*, 1998 ; Glock & Kleen, 2020.

³⁵ Boyce & Herd, 2003.

³⁶ Biernat, Vescio & Manis, *et al.*, 1998.

³⁷ Personal communication with Professor Monica Biernat, October 2011.

These bodies of research refine our understanding of the conditions under which stereotyping is activated, and the impact that activation can have on individuals in groups, as well as the group's performance. Regardless of the source of stereotyping, we have come to understand that the presence of stereotyping has the potential to impact teams, but the conditions under which that happens are inconclusive.³⁸ This lack of convergence on the specific conditions under which stereotyping affects team performance has less to do with the nature of stereotyping than it does the many complicated ways that diverse teams can and are studied. To date, much of the research has focused on isolating singular aspects of diversity and then examining those with respect to proximal measures of team performance. It has been suggested that research on stereotyping in groups needs to be rededicated to understanding diversity as a *combination* (emphasis added) of different dimensions of differentiation, rather than continuing to focus on singular aspects of diversity.³⁹ The current research leverages the shifting standards model to detect and measure a variety of sources of stereotyping on teams, and then combines them to determine the overall presence of stereotyping and its collective impact on team performance.

Despite the copious attention paid to stereotyping and team performance, there are relatively few studies that attempt to measure the level of stereotyping present on a team and its impact on *objectively measured* team performance outcomes. We suggest that there is an empirical gap in the area that connects the absolute value of the level of stereotyping on a team to that team's objective performance measures. This study addresses that empirical gap by utilizing the shifting standards model in conjunction with ML-SEM to measure the level of stereotyping in several teams and then leveraging juried competition standards to measure the teams' performance.⁴⁰ In so doing, we were able to determine the extent to which the presence and level of stereotype-based judgments impact a team's objective performance.

Specifically, in this research we sought to answer whether or not there was evidence of the shifting standards effect at the team level in a sample of 37 military teams competing head to head in a military skills competition, and if so, whether or not there was variability in the shifting standards index across the teams. If there was evidence of variability in the shifting standards effect across these teams, then we sought to determine what specific sources of stereotyping were responsible for that variability. In particular, and as reviewed in the literature mentioned, we examined variability in the shifting standards index associated with gender, race, and cadet class as well as an assimilation of all those categories together. Finally, we sought to determine if there was a relationship between the level of overall stereotyping on each team (as indicated by that team's shifting standards index) and objective measures of team performance.

³⁸ Millikin & Martens, 1996 ; Cohen & Bailey, 1997.

³⁹ Van Knippenberg & Schippers, 2007 ; Clauss-Ehlers *et al.*, 2019.

⁴⁰ Hair & Sarstedt, 2019.

Research Questions

In this study we examined the impact of stereotyping on team performance utilizing the shifting standards model and ML-SEM. In doing so, our goal was to answer the following research questions:

1. Is there evidence of gender-based stereotyping across these teams?
2. Is there evidence of race-based stereotyping across these teams?
3. Is there evidence of cadet class-based stereotyping across these teams?
4. Is there a relationship between the presence of stereotyping (as evidenced by the presence of shifting standards) and team performance in the competition?

Methods

The study consisted of 37, nine- to 13-person teams that competed in a military skills competition held at the United States Military Academy (USMA) in West Point, New York. By regulation, each team was comprised of participants from each of the four classes at West Point (freshman, sophomore, junior, and senior) and no less than one woman. There was no race-based team composition requirement. A total of 447 cadets participated in the competition. Of these, 18% were women, 24% were non-White, and 62% were lower class (freshmen and sophomore) cadets. No team had more than two women members, and six teams had only one woman. There were three all-White teams.

Competition Overview

The Sandhurst Competition is held at the United States Military Academy every April. The competition is designed to enhance small unit leadership qualities, develop teamwork, and promote military excellence. It consists of a series of knowledge- and skill-based military events that are timed and judged over the period of three days. Each year, teams from all over the world participate in the three-day competition study:

Table 1: List and Description of Sandhurst Competition Events

Event	Event Description
Marksmanship	Weapons firing from multiple positions, direct fire planning, and fire control techniques
Land Navigation	Complete a 5-mile orienteering course
Obstacle Course	Negotiate a course with 10 challenging obstacles
Weapons and Grenades	Negotiate a live hand grenade course; disassemble and assemble a variety of weapons systems to include the M4, M240, M240B, M9, and foreign weapons
Rappel	Negotiate and descend a 75-foot rock face by constructing the team's own Swiss seats and equipment slings

Zodiac Boat Race	Launch and paddle a ¼ mile water course
Rope Bridge	Conduct a stream crossing by utilizing an improvised bridge constructed out of the team’s provided equipment
Tactical Combat Casualty Care	Perform lifesaving battlefield first aid after encountering an Improvised Explosive Device; conduct combat air evacuation techniques
Department of Military Instruction Challenge	Demonstrate lateral thought, mental agility, ability to assimilate a problem, and decision-making under physical stress
Time on Course	The time, including penalties, it takes the team to complete the entire competition

Measures

Survey Instrument

We developed an online instrument that contained leadership competencies utilized by the United States Military Academy’s Cadet Evaluation Report (CER). When this research was being conducted the CER was a developmental assessment instrument familiar to all cadets and cadet cadre and was comprised of nine different leadership competencies :

Table 2: List and Description of Leadership Competencies

Competency	Description
Communicating	Displays good oral, written, and listening skills for individuals/groups
Decision-making	Employs sound judgment, logical reasoning, and uses resources wisely
Motivating	Inspires, motivates, and guides others toward mission accomplishment
Planning	Develops detailed, executable plans that are feasible, acceptable, and suitable
Executing	Shows tactical proficiency, meets mission standards, and takes care of people/ resources
Assessing	Uses after-action and evaluation tools to facilitate consistent improvement
Developing	Invests adequate time and effort to develop individual subordinates as leaders
Building	Spends time and resources improving teams, groups, and units
Learning	Seeks self-improvement and organizational growth; envisioning, adapting, and leading change

For this research, the survey instrument asked cadets to both rate and rank each other and themselves in each of the nine leadership competencies. The ratings were scored

on a 6-point Likert scale (Very Ineffective, Ineffective, Somewhat Ineffective, Somewhat Effective, Effective, Very Effective), and the rankings required each cadet to list each of their teammates in order of their effectiveness in each competency. The first position was reserved for the cadet judged to be the most effective in that competency, and the final position was reserved for the cadet judged to be the least effective in that competency. See Appendix A for a description of the cadet-level variables and team-level variables.

Procedure

After training together for six weeks, but eight weeks before the competition took place, cadets completed an online survey in which they rated and ranked their teammates and themselves in each of nine dimensions of a leadership competency model outlined by the USMA Cadet Evaluation Report. At the end of the competition, cadets were again asked to rate and rank their teammates and themselves in those same nine dimensions. This allowed us to compute the level of stereotyping that occurred within teams. Team performance outcomes were recorded at the completion of the competition.

Results

We sought to determine if stereotyping was occurring on these military teams as indicated by the shifting standards effect. We also wanted to know if the stereotyping was variable across the teams and, if so, whether that variability was related to variability in team performance. Specifically, we were interested in what association, if any, existed between levels of stereotyping, as indicated by the shifting standards effect, and team performance.

We used multilevel structural equation modelling (ML-SEM) to determine the effects of stereotyping on team performance. Level 1 is the Cadet Level, and Level 2 is the Team Level. We fit two models, Model A and Model B, to answer Research Questions 1 and 2. We fit two models, Model C and Model D, to answer Research Question 3. See Appendix B for a thorough statistical description of these four models.

A total of 313 of a possible 447 cadets on 37 teams responded to the survey. This was a 70% response rate overall. Teams ranged in size from nine to 13 cadets. The average response rate per team was 7.5 cadets. The minimum number of respondents for a team was two cadets, and the maximum was 12 cadets. Table 3 provides a descriptive for the composition of the competition participants with respect to gender, race, and cadet class.

Table 3: Participants by Gender, Race, and Cadet Class.

	Male (82%)		Female (18%)		Total
	White (80%)	Non-White (20%)	White (56%)	Non-White (44%)	
Upper Class	122 (27%)	31 (7%)	14 (3%)	5 (1%)	172 (38%)
Lower Class	172 (38%)	44 (10%)	30 (7%)	29 (6%)	275 (62%)
Total	294	75	44	34	447 (100%)

Not all cadets on all teams participated in the survey ; however, because all teams participated with at least two raters, every cadet in the competition was rated and ranked. There was no significant difference between respondents and non-respondents with respect to gender and race. Lower class cadets responded at slightly higher rates (73% vs. 75%). There was no significant difference in shifting standards between respondents and non-respondents.

Table 4 (below) shows the descriptive statistics at the team level for demographics, size, and response rate. Of note, teams have much more variability with respect to the number of lower class cadets (3 to 11) and non-White cadets (0 to 7) compared to females (1 to 2). Team response rate was not significantly associated with team composition, team-level shifting standards with respect to gender, race, or class, or team performance. Table 5 (further down on this page) contains the bivariate relationship between the Shifting Standards Index and Gender, Race, and Cadet Class.

Table 4 : Team Level Descriptive Statistics

	Mean	Standard Deviation	Minimum	Maximum
Response rate	7.486	2.854	2	12
Team size	11.811	0.569	10	13
Females	1.838	0.374	1	2
Males	9.973	0.440	9	11
Lower Cadet Class	7.162	1.692	3	11
Upper Cadet Class	4.649	1.531	3	7
Non-White	2.676	1.796	0	7
White	9.135	1.719	5	12

Table 5 : Shifting Standards Index for Each Demographic Category

Shifting Standards Index						
Status characteristic	Observations	Mean	Standard deviation	Minimum	Maximum	Mean differences
Female	68	0.168	0.820	-1.166	2.699	0.196*
Male	368	-0.035	0.398	-1.218	1.897	
Non-White	99	0.024	0.550	-1.000	2.490	0.040
White	337	-0.013	0.474	-1.220	2.700	
Lower Cadet Class	264	0.054	0.493	-1.220	2.700	0.157**
Upper Cadet Class	172	-0.093	0.480	-1.140	2.489	

* $p < 0.05$, ** $p = 0.001$

Gender

We expected female cadets to have lower status than male cadets and therefore generate higher levels of the shifting standards index than male cadets. In this sample, females have significantly higher shifting standards index (SSI) than do males (+0.168 vs. -0.036). Their standard deviation is larger, and they have more variability around the mean as well as larger outliers than do males.

Race

We expected non-White cadets to have lower status than White cadets and, therefore, generate higher levels of the shifting standards index than White cadets. However, we found that White and non-White cadets do not differ appreciably with respect to SSI.

Cadet Class

In accordance with institutional norms, lower class cadets have lower status than upper class cadets, and so we expected to see a relationship between cadet class and SSI. In this sample, lower class cadets have significantly higher SSI, on average, indicating they generate more variability in the shifting standards index than upper class cadets.

We fit Model A to determine whether there was evidence of shifting standards on teams with respect to a cadet’s diversity level, as indicated by their diversity index (DVI) and their shifting standards index (i.e., how much stereotyping that cadet generated in others). On average, across all teams, there is evidence that a cadet’s level of diversity is indeed positively associated with SSI (Est. = 0.117, SE = 0.037, $p = 0.001$), as shown in Table 6 (below). The more that cadets possessed lower status characteristics (as indicated by their DVI), the more evidence we found of stereotyping, as indexed by SSI.

Table 6 : Model Parameters Predicting Shifting Standards Indices (SSI) by Diversity Index and Cadet Gender, Race, and Class (n = 447)

Model	Random coefficient	Mean (β)			Variance (σ^2)		
		Est. (α)	SE	p -value	Est.	SE	p -value
A	β_1 SSI on DVI	0.117	0.037	0.001	0.000	0.079	0.997
	β_2 SSI on female	0.176	0.094	0.061	0.148	0.045	0.001
B	β_3 SSI on race	0.004	0.049	0.934	0.001	0.040	0.972
	β_4 SSI on class	0.143	0.052	0.006	0.005	0.021	0.802

Note. The results of a global hypothesis test to determine the impact of all predictor variables simultaneously were significant ($\chi^2 = 35.00$, $df = 3$, $p < 0.001$), that $\alpha_2 = \alpha_3 = \alpha_4 = 0$, rejecting the null.

For Model B, we tested all relevant two-way and three-way interactions with the main predictors (female, race, and class) and found that there was no evidence of either two-way or three-way interactions for female, race, and class with respect to SSI. The likelihood ratio test (adj. LRTS = 1.92, $df = 4$, $p = 0.75$) confirmed this.

We fit Model B to determine whether there was evidence of shifting standards on teams with respect to gender, race, and cadet class. The highest level of shifting standards was associated with female cadets (Est. = 0.176, SE = 0.094, $p = 0.061$) and lower-class cadets (Est. = 0.143, SE = 0.052, $p = 0.006$). Non-White cadets did not generate an appreciable level of the shifting standards effect (Est. = 0.004, SE = 0.049, $p = 0.934$), suggesting that race-based stereotyping was not statistically significantly activated in this context.

To determine if there was variability in shifting standards across teams, we examined Model B and found that teams did vary with respect to their level of stereotyping, but interestingly, only with respect to gender (Est. = 0.148, SE = 0.040, $p < 0.001$; see Table 6, above). The variance associated with race stereotyping (Est. = 0.001, SE = 0.04, $p = 0.972$) and class stereotyping (Est. = 0.005, SE = 0.021, $p = 0.802$) was striking both in its modesty and in its similarity. This suggests a similar or shared understanding or perception of lower-class and non-White cadets across all teams. Teams demonstrate a general lack of stereotyping of non-White cadets similarly across all 37 teams.

To determine if there is a relationship between team-level shifting standards indices and the team's performance, we fit Model C and Model D regressing SSI on gender, race, cadet class, and DVI at Level 1 and regressing team performance on the team-level shifting standards at Level 2 to determine the team-level effects. We found that there was no relationship between a team's level of shifting standards and its overall performance in the competition, as indicated by number of points earned by the team.⁴¹ We then looked at team performance in each of the 10 events in the competition to determine if there was a relationship between a team's level of shifting standards and its event-specific performance. See Table 7 (next page) for the individual events in Model C. To counteract the problem of inflated Type I error rate due to multiple comparisons across events, we used the Bonferroni corrected α -level of 0.005.

We found that indeed, in a few instances, a team's level of shifting standards, either as a whole or with respect to gender or race, was associated with its performance. For example, a team's overall level of shifting standards (i.e., its global level of stereotyping) was associated with a team's performance in the marksmanship event (weapon firing accuracy) such that higher levels of shifting standards were associated with lower levels of team performance in that event (Est. = -25.128, SE = 8.336, $p = 0.003$, Std. Est. = -0.829). Similarly, a team's overall level of shifting standards was associated with a team's performance in the DMI Challenge event (see Table 1 for a description) such that higher levels of shifting standards were associated with lower levels of team performance in that event (Est. = -57.798, SE = 2.320, $p = 0.000$, Std. Est. = -0.099). Conversely, in the rappelling event (rappelling the entire team and its equipment down a 75-foot cliff), higher

⁴¹ See top row of Table 7, next page.

Table 7 : Model Parameters Predicting Team Performance in 10 Events by Team-Level Shifting Standards Indices (SSI) by Cadet Gender, Race, and Class, and Diversity Index (n = 447).

Event	Model C				Model D				Model D				Model D			
	β1 SSI on DVI				β2 SSI on Female				β3 SSI on Race				β4 SSI on Class			
	Est.	SE	p-value	stde4	Est.	SE	p-value	stde1	Est.	SE	p-value	stde2	Est.	SE	p-value	stde3
Total points	3.040	118.060	0.979	0.055	-0.434	0.715	0.544	-0.167	11.172	18.669	0.550	0.517	-1.678	11.370	0.883	-0.127
Marksmanship	-25.128	8.336	0.003*	-0.829	-0.881	0.603	0.144	-0.341	-11.208	7.694	0.145	-0.616	-4.661	7.023	0.507	-0.380
Land Navigation	-0.707	96.611	0.994	-0.013	-0.728	0.546	0.182	-0.280	10.804	12.043	0.377	0.499	-0.529	8.445	0.950	-0.040
Obstacle Course	-2.403	140.776	0.986	-0.044	0.154	0.906	0.863	0.059	-2.607	65.158	0.969	-0.098	-6.765	17.821	0.704	-0.527
Rope Bridge	27.090	41.362	0.513	0.650	0.193	0.824	0.816	0.075	12.627	8.185	0.123	0.754	2.500	10.322	0.809	0.183
Rappelling	21.566	4.955	0.000**	0.894	0.345	0.541	0.524	0.156	8.570	2.398	0.000**	0.808	4.968	3.771	0.188	0.471
TCCC	27.460	26.135	0.293	0.688	0.367	0.616	0.551	0.143	7.461	24.400	0.760	0.296	7.687	9.400	0.413	0.665
Weapons	0.280	105.090	0.998	0.005	-0.257	0.600	0.668	-0.104	6.841	12.215	0.579	0.311	-7.668	5.249	0.144	-0.827
Boat Race	-26.464	19.042	0.165	-0.797	-0.157	0.672	0.815	-0.062	-10.020	2.288	0.000**	-0.900	1.861	10.488	0.859	0.121
DMI Challenge	-57.798	2.320	0.000**	-0.099	-1.454	0.530	0.006*	-0.574	8.113	1.952	0.000**	0.655	-4.091	4.301	0.342	-0.339
Course time	11.099	114.564	0.923	0.207	0.094	0.814	0.908	0.036	-9.357	32.590	0.774	-0.397	-1.465	14.211	0.918	-0.108

levels of shifting standards were associated with higher levels of team performance (Est. = 21.566, SE = 4.955, $p < 0.001$, Std. Est. = 0.894). When aggregated for the entire competition, these effects are not apparent.

Additionally, a team's level of shifting standards with respect to specific sources of stereotyping (gender and race) had associations with team performance in some events. For example, in one instance, higher levels of shifting standards with respect to female cadets was associated with lower levels of team performance in the DMI Challenge event (Est. = -1.454, SE = 0.530, $p = 0.006$, Std. Est. = 0.006). Also, higher levels of shifting standards with respect to race was associated with higher levels of team performance in the rappelling event (Est. = 8.570, SE = 2.398, $p < 0.00$, STDE = 0.808), and the DMI Challenge event (Est. = 8.113, SE = 1.952, $p < 0.00$, STDE = 0.655), but was associated with lower levels of team performance in the Zodiac boat course (Est. = -10.020, SE = 2.288, $p < 0.00$, STDE = -0.900).

Discussion

The purpose of this study was to examine how multiple sources of stereotyping can impact team performance. We found no evidence that global stereotyping (i.e., stereotyping with respect to gender, race, and class altogether) was associated with overall team performance, as measured by either points earned or placement with respect to other teams in a 10-event competition. It has been argued that in field settings, the impact on team performance of stereotyping with respect to demographic variables may be obviated by functional team norms such as collectivism and teamwork (Bell, 2007). Given the nature of these tightly knit teams, and the pressure of the competition, it may be that the superordinate goal of winning, indeed overshadowed any impacts stereotyping might have had on team performance. Additionally, another perspective suggests that team training specifically focused on task-relevant knowledge, skills, and abilities has a moderate effect on increasing adaptive team mechanisms, which result in higher levels of team cohesion, motivation, and performance (Mathieu *et al.*, 2008). Teams in this study trained together for several months prior to the competition, potentially creating the conditions that reduced the effects of stereotyping on team performance. Establishing strong and functional team norms through effective team training mechanisms may mitigate the potentially deleterious effects of stereotyping on team performance.

Although global team-level stereotyping was not associated with a reduction in team performance overall, the impact of stereotyping on the Sandhurst teams was more complicated and nuanced. Specifically, we found evidence that team-level stereotyping influenced three specific events. Higher levels of team-level stereotyping were associated with lower levels of performance in both the marksmanship and the DMI Challenge events, and higher levels of performance in the rappelling event.

Additionally, stereotype threat research suggests that the threat of confirming a negative stereotype about a group one belongs to can actually create enough performance pressure to interfere with one's performance.⁴² Given that pooled task interdependence makes individual contributions to the team more visible, stereotyped individuals may have felt more vulnerable to their group's stereotype and therefore experienced some fear of fulfilling that stereotype. In response, they may have performed more poorly on the task, thereby reducing the team's overall performance in the event.

The second event in which team-level global stereotyping had a deleterious effect on team performance was the DMI Challenge. This event is the last event of the three-day series. Teams are scored in two activities in the DMI Challenge. The first activity challenges the team to figure out how to disassemble a 155 mm howitzer cannon from its firing position, prepare it for movement, load its auxiliary equipment onto its base plate assembly, and then, without assistance, move the entire cannon and its load 100 yards across an open field. This task requires high levels of coordination, cooperation, knowledge sharing, and strength. The second task in the DMI Challenge is an intellectually challenging task requiring teams to answer a series of questions about what they saw as they travelled through the many different sites of the competition's events. This task tests cadets' level of attention to detail and requires them to assemble subtle bits of information they should have noticed while they were competing. For example, the test posed questions such as "At the Weapons event, there was an enemy armoured personnel carrier (APC) parked on the right side of the medical tent. What nation's flag was painted on the side of that APC?". These questions were designed to challenge the cadets' attention to detail and memory under pressure. The cadets were then tested on how well they could construct consistent responses from their shared mental models. Success in the DMI Challenge depends on physical strength, teamwork, information sharing, creative problem solving, mental agility, and the team's ability to tap into its collective team memory. Some research suggests that in diverse groups, stereotyping interacts with information sharing and decision-making in a way that disrupts a team's ability to process task-relevant perspectives and information.⁴³ The teams that generated higher levels of global stereotyping may have communicated and shared information less than teams that had lower levels of stereotyping. This may have impeded the level of cooperation among teammates while solving both challenging problems (to move the howitzer and to reconstruct the factual details along the course). Presumably this could have reduced their ability to tap into the complementary talents and ideas of individual cadets, thereby compromising the team's performance. Future studies could address this possibility.

⁴² Steele & Aronson, 1995 ; Shih *et al.*, 1999 ; Rosenthal *et al.*, 2007 ; Pennington *et al.*, 2016.

⁴³ Van Knippenberg *et al.*, 2004.

In contrast to the association that team-level stereotyping had with team performance at the marksmanship range and in the DMI Challenge, higher levels of team-level stereotyping was associated with higher levels of performance in the rappelling event. Although counterintuitive, this effect might be explained by how the use of some stereotypes may improve a team's ability to leverage its members' varied abilities. For example, in the rappelling competition, part of the team's success is getting several strong team members to the bottom of the cliff first, so they can act as belay for the rest of the team. Performance at belay is enhanced by height and exceptional upper and lower body strength. These demands may have influenced teams to send their tallest and most physically fit cadets down the cliff face first, to take the role of belay. Given these selection standards, teams may have filtered their choices by gender and race to select team members to be on belay. This might have enabled the rest of the team to rappel more rapidly, and therefore score higher in the event.

In a similar vein, the rappelling event requires sequential interdependence and, as such, the determination for who goes over the cliff first is an important one. All the subsequent tasks depend on the one before it. If stereotyping leads to the team picking the larger, more athletic looking males to rappel first and then set up belay for the rest of the team and its equipment, then this perhaps generates a positive outcome for the team's performance. In this way, a higher level of stereotyping may have been associated with better team performance. Again, this may be a focus for further research.

We were also interested in examining the ways that individual sources of stereotyping with respect to gender, race, and cadet class impacted (or not) the team's performance in individual events of the Sandhurst Competition. The existence of higher levels of shifting standards when female cadets are being judged is not surprising as women are stereotype-relevant targets in this uniquely masculinized environment of military skills competition.⁴⁴ However, stereotyping of female cadets was associated with lower team performance in only one event, the DMI Challenge event.

Also, with regard to the team's inability to tap into all of its talent, research on fault lines lends some insight into the impact of demographic diversity on team performance.⁴⁵ Fault lines occur in teams when members fall into two or more distinct demographic categories (in this case, female cadets and male cadets, or White cadets and non-White cadets). "Fault lines" are fractures or divides that lay between these demographic categories in teams and can be activated by external forces (such as competition pressures). Fault lines can be responsible for conflict that impacts team performance when the demographic categories are related to the group's task.⁴⁶ Additionally, fault line research suggests that in teams tasked

⁴⁴ Mettrick & Cowan, 1996 ; Boldry *et al.*, 2001 ; Oparah, 2012.

⁴⁵ Mannix & Neale, 2005 ; Thatcher & Patel, 2011 ; Pham *et al.*, 2012.

⁴⁶ Lau & Murnighan, 1998 ; Lau & Murnighan, 2005.

with creative thinking or sharing mental models (not unlike the DMI Challenge event), gender fault lines can affect team performance negatively.⁴⁷ This may explain the effects associated with gender stereotyping and team performance in this specific event.

Most striking were the relatively low levels of racial bias as indicated by the absence of any substantial shifting standards effect with respect to White and non-White cadets. This may be attributed to the strong prescriptive norms at the US Military Academy mandated by the Character programme,⁴⁸ an institutionalized four-year developmental programme designed to improve interracial understanding, dialogue, and respect. Although it includes gender as a topic, what might retard the impact of that aspect of the programme is the undeniably masculine nature of the Military Academy. This finding is consistent with other research related to race and military teams, which found similar patterns of relatively insignificant effects with respect to racial stereotype activation using the shifting standards model in military settings.⁴⁹

The nature of military teams also may contain additional explanatory insights into why stereotyping did not have a relationship with overall performance in the Sandhurst competition. More than most, these participants have engaged in formative experiences that disconfirm traditional stereotypes. Given the nature of the United States Military Academy and cadet life, Allport's (1954) contact hypothesis comes to mind, potentially, as an explanatory concept. This hypothesis suggests that when majority group members have meaningful and interdependent interactions with stereotyped individuals, stereotyping is reduced. The implication is that if the aim is to reduce intergroup conflict, contact and interaction between majority and minority group members should be increased. Rothbart and John (1985) refined the conditions under which true stereotype reduction would take place, and these are all relevant to this study. They posited that to cognitively reduce stereotyping, majority individuals must be susceptible to acknowledging the existence of stereotype disconfirming data from minority individuals, and that they then must generalize those impressions to the minority's group. In addition, majority and minority members must interact in a variety of settings and be focused on a common goal. When all these conditions are met, stereotyping is reduced.⁵⁰ At the US Military Academy, where there is constant contact between highly talented members of all ethnicities and genders, much of that contact is in different settings and often focused on superordinate goals. Given that these conditions meet the requirements for stereotype reduction established by Rothbart and John (1985), it is no wonder that stereotyping's impact on team performance is isolated and minimal.

⁴⁷ Pearsall *et al.*, 2008.

⁴⁸ Cf. <https://www.westpoint.edu/character-program>, accessed September 30, 2020.

⁴⁹ Biernat, Crandall *et al.*, 1998.

⁵⁰ Rothbart & John, 1985 ; Pettigrew & Tropp, 2006 ; Pettigrew & Tropp, 2008.

The implications of this study are relevant to both research and practice. With respect to practice, these findings might provide insights for other domains, such as education and business. As the use of teams expands in education and in business, it is imperative that we better understand the conditions that optimize team performance. For leaders and executives, this may suggest a change in behaviours. Specifically, in organizations, leaders tend to put lower status people in situations that require pooled interdependence. This gives them an opportunity to demonstrate their individual talents, and therefore provides them an opportunity to shine on their own, rather than be dependent on others. However, this research suggests if the goal is to optimize team performance and denigrate the potential deleterious impact of stereotyping in teams, that in addition to providing lower status individuals opportunities to demonstrate their expertise, it should be ensured they are part of mixed teams that face emergent challenges that demand reciprocal interdependence. So, rather than giving low status individuals autonomy, independence, and asking them to “prove” themselves as solo artists, they should be given more chances to collaborate with team mates on challenging tasks.

In the future, it would be beneficial to examine the perceived relative masculinity and femininity of each of the leadership competencies in which cadets rated each other. It would be informative to see if there is a gender- or race-based association between how masculine or feminine a competency is perceived to be and who is rated and ranked higher in those competencies. This would lend deeper insight into the way gender-, race-, and class-based stereotyping impact interpersonal assessments of competence.

Research that extends this work could more closely look at specific sources of stereotyping (gender, race, and cadet class) and their relationship to a variety of leadership competencies. Specifically, it would be useful to study perceptions of the relative masculinity and femininity of leadership competencies and how they are rated with respect to their importance to the overall construct of leadership effectiveness. It would also help to look at the perceived relative masculinity and femininity of specific tasks in different professions and then compare that to the gender of those individuals judged as the most effective in those tasks.

In addition, researchers could shift to examining the ways that team members’ *liking* for each other influences performance. Research toward those ends could engage the work of Todd Pittinsky on allophilia, which focuses on the ways that positive feelings towards out-group members enhance team process and performance.⁵¹ The ultimate purpose of this stream of research in stereotyping and team performance is to better understand the myriad ways that stereotyping can impact team process and performance so that education and training programmes can be designed to increase the positive effects and decrease the deleterious effects of these dynamics for leaders and members of teams.

⁵¹ Shih *et al.*, 1999 ; Pittinsky, 2005 ; Pittinsky *et al.*, 2011.

Limitations

There are several limitations of this research. Although the aim here was to more directly link measures of team-level stereotyping and team performance, there are many mediators whose measurement would lend more nuanced interpretations of the effects of stereotype-based judgments on team performance. For example, Chatman and Flynn (2001) found that cooperative group norms reversed some of the deleterious effects demographic heterogeneity had on group outcomes. In future research, scholars could examine group process-oriented mediators to determine which has the potential to obviate the effects of stereotyping and prejudice.

In addition, this research may not lend insight into other kinds of teams, such as virtual or geographically distributed teams, whose potential for contact-based stereotyping is somewhat limited. Future research might target such teams to determine which, if any, sources of stereotyping prevail, when physical presence is not a dominant aspect of interpersonal contact. Furthermore, in the absence of individual-level performance data for each event in the competition, as well as observations of individuals with respect to the competency domains, it is more difficult to determine just how biased or accurate some interpersonal assessments are. It has been argued that, to its own detriment, field research on team performance far too often ignores the influence of task-specific skill diversity (Jackson *et al.*, 2003). In the future, we suggest designing the study to capture the actual task-specific skills of each of the cadets on the teams, thereby allowing for more accurate inferences regarding the interpersonal assessments of competence rendered by each cadet.

With respect to research, we strongly recommend a more multimodal approach to this kind of study. This would improve its capacity to offer insights into the phenomena of interpersonal biases and how they impact team performance. Specifically, the use of qualitative methods, such as ethnography, could better capture the lived experience and perceptions of the cadets going through the training and the competition. Giving voice to participants provides the opportunity to garner a much more detailed understanding of the dynamics of team life. This more robustly qualitative approach could provide more insight into the multidimensional nature of diversity, and hence how it is interpreted in this study (Jackson *et al.*, 2003).

Conclusion

This research adds to the literature on stereotyping and team performance by leveraging the strength of multilevel structural equation modelling to enhance the ability of the shifting standards model to detect and measure stereotyping at the team level. Hence, this research was able to provide insight into the multidimensionality of diversity's effect on team performance and suggests that the nature of the specific tasks faced by teams might have an impact on the ways stereotyping manifests in the team's performance. The nature of the

interdependence of a task may have had a moderating effect on the impact stereotyping has on the team's performance in that task, and this opens the door for further investigation of high-performance teams.

This research complicates the claims made about diversity and team performance and challenges future researchers to look more closely at several moderating variables. It is not just that diversity makes groups more effective, and it is not just that diversity activates status-based classifications that can undermine the benefits of diversity. The present research and methodology suggest that there are other factors at play, which are important to examine to truly understand stereotyping in the context of teams.

References

- ADDISON, R. & C. HAIG**, *Performance Architecture, Care and Feeding of a High-Performance Team*, A BPT Trends Column, 2012, pp.1-5 : <https://www.bptrends.com/bpt/wp-content/publicationfiles/12-04-2012-PA-Care%20and%20Feeding-Addison%20and%20Haig-final.pdf>.
- ALLPORT, Gordon W.**; *The Nature of Prejudice*, Reading, MA, Addison-Wesley, 1979.
- ANCONA, D.G. & D.F. CALDWELL**, "Demography and Design : Predictors of New Product Team Performance", *Organization Science*, vol.3, n°3, 1992, pp.321-341 : doi.org/10.1287/orsc.3.3.321.
- ANTINO, M., R. RICO & S.M. THATCHER**, "Structuring Reality through the Faultlines Lens: The Effects of Structure, Fairness, and Status Conflict on the Activated Faultlines – Performance Relationship", *Academy of Management Journal*, vol.62, n°5, 2019, pp.1444-1470 : doi.org/10.5465/amj.2017.0054.
- AUSTIN, J.T., P. VILLANOVA, R. KATZELL, L.G. HUMPHREYS, A. RYAN, R. HENE & K. BOYLE**, "The Criterion Problem : 1917-1992", *Journal of Applied Psychology*, vol.77, n°6, 1992, pp.836-874 : doi.org/10.1037/0021-9010.77.6.836
- AUSTIN, J.R.**, "Transactive Memory in Organizational Groups : The Effects of Content, Consensus, Specialization, and Accuracy on Group Performance", *Journal of Applied Psychology*, vol.88, 2003, pp.866-878.
- BACHRACH, D.G., K. LEWIS, Y. KIM, P.C. PATEL, M.C. CAMPION & S. THATCHER**, "Transactive Memory Systems in Context: A Meta-Analytic Examination of Contextual Factors in Transactive Memory Systems Development and Team Performance", *Journal of Applied Psychology*, vol.104, n°3, 2019, pp.464-493 : doi.org/10.1037/apl0000329.
- BANAJI, M.R. & C.D. HARDIN**, "Automatic Stereotyping", *Psychological Science*, vol.7, n°3, 1996, pp.136-141. doi.org/10.1111/j.1467-9280.1996.tb00346.x
- BELL, S.T.**, "Deep-level Composition Variables as Predictors of Team Performance : A Meta-Analysis", *Journal of Applied Psychology*, vol.92, n°3, 2007, pp.595-615 : psycnet.apa.org/doi/10.1037/0021-9010.92.3.595.
- BERGER, J., B.P. COHEN & M. ZELDITCH**, "Status Characteristics and Social Interaction", *American Sociological Review*, vol.37, n°3, 1972, pp.241-255 : doi.org/10.2307/2093465.
- BERGER, J., R.Z. NORMAN, J.W. BALKWELL & R.F. SMITH**, "Status Inconsistency in Task Situations : Test of Four Status Processing Principles", *American Sociological Review*, vol.57, n°6, 1992, pp.843-855 : <https://doi.org/10.2307/2096127>.
- BERGER, J., C.L. RIDGEWAY, M.H. FISEK & R.Z. NORMAN**, "The Legitimation and Delegitimation of Power and Prestige Orders", *American Sociological Review*, vol.63, n°3, 1998, pp.379-405 : <http://dx.doi.org/10.2307/2657555>.

BIERNAT, M. & M. MANIS, “Shifting Standards and Stereotype-Based Judgments”, *Journal of Personality & Social Psychology*, vol.66, n°1, 1994, pp.5-20: <http://www.ncbi.nlm.nih.gov/pubmed/8126651>.

BIERNAT, M., C.S. CRANDALL, L.V. YOUNG, D. KOBRYNOWICZ & S.M. HALPIN, “All That You Can Be: Stereotyping of Self and Others in a Military Context”, *Journal of Personality & Social Psychology*, vol.75, n°2, 1998, pp.301-17 : <http://www.ncbi.nlm.nih.gov/pubmed/9731310>.

BIERNAT, M., M.J. TOCCI & J.C. WILLIAMS, “The Language of Performance Evaluations: Gender-Based Shifts in Content and Consistency of Judgment”, *Social Psychological & Personality Science*, vol.3, n°2, 2011, pp.186-192 : <https://doi.org/10.1177%2F1948550611415693>.

BIERNAT, M., M. MANIS & T.E. NELSON, “Stereotypes and Standards of Judgment”, *Journal of Personality & Social Psychology*, vol.60, n°4, 1991, pp.485-499 : [psycnet.apa.org/doi/10.1037/0022-3514.60.4.485](https://doi.org/10.1037/0022-3514.60.4.485).

BIERNAT, M. & T.K. VESCIO, “Categorization and Stereotyping : Effects of Group Context on Memory and Social Judgment”, *Journal of Experimental Social Psychology*, vol.29, n°2, 1993, pp.166-202 : <https://doi.org/10.1006/jesp.1993.1008>.

BIERNAT, M., T.K. VESCIO & M. MANIS, “Judging and Behaving toward Members of Stereotyped Groups: A Shifting Standards Perspective”, pp.151-175 in C. Sedikides, J. Schopler & C.A. Insko (eds.), *Intergroup Cognition and Intergroup Behavior*, Mahwah, NJ, Lawrence Erlbaum, 1998.

BIERNAT, M. & D. KOBRYNOWICZ, “Gender- and Race-Based Standards of Competence : Lower Minimum Standards but Higher Ability Standards for Devalued Groups”, *Journal of Personality & Social Psychology*, vol.72, n°3, 1997, pp.544-557 : <https://doi.org/10.1037//0022-3514.72.3.544>.

BODENHAUSEN, G.V., “Diversity in the Person, Diversity in the Group: Challenges of Identity Complexity for Social Perception and Social Interaction”, *European Journal of Social Psychology*, vol.40, n°1, 2010, pp.1-16.

BOLDRY, J., W. WOOD & D.A. KASHY, “Gender Stereotypes and the Evaluation of Men and Women in Military Training”, *Journal of Social Issues*, vol.57, n°4, 2001, pp.689-705: [psycnet.apa.org/doi/10.1111/0022-4537.00236](https://doi.org/10.1111/0022-4537.00236).

BOYCE, L.A. & A.M. HERD, “The Relationship between Gender Role Stereotypes and Requisite Military Leadership Characteristics”, *Sex Roles*, vol.49, n°7-8, 2003, pp.365-378.

BREWER, M.B. & K.P. PIERCE, “Social Identity Complexity and Outgroup Tolerance”, *Personality & Social Psychology Bulletin*, vol.31, n°3, 2005, pp.428-37: <https://doi.org/10.1177%2F0146167204271710>.

CHATMAN, J.A. & F.J. FLYNN, “The Influence of Demographic Heterogeneity on the Emergence and Consequences of Cooperative Norms in Work Teams”, *Academy of Management Journal*, vol.44, n°5, 2001, pp.956-974 : <https://doi.org/10.5465/3069440>.

CLAUSS-EHLERS, C.S., D.A. CHIRIBOGA, S.J. HUNTER, G. ROYSIRCAR & P. TUMMALA-NARRA, “APA Multicultural Guidelines Executive Summary : Ecological Approach to Context, Identity and Intersectionality”, *Am. Psychologist*, vol.74, n°2, 2019, pp.232-244: <https://doi.org/10.1037/amp0000382>.

COLLINS, P.H. & S. BILGE, *Intersectionality*, Hoboken, NJ, John Wiley & Sons, 2020.

CRAWFORD, E.R. & J.A. LEPINE, “A Configural Theory of Team Processes: Accounting for the Structure of Taskwork and Teamwork”, *Academy of Management Review*, vol.38, n°1, 2013, pp.32-48 : <https://doi.org/10.5465/amr.2011.0206>.

D’ACUNTO, F., U. MALMENDIER & M. WEBER, *Gender Roles and the Gender Expectations Gap*, Working Paper 26837, National Bureau of Economic Research, 2020: <https://dx.doi.org/10.2139/ssrn.3547537>.

DEVARAJ, S. & K. JIANG, “It’s about Time – A Longitudinal Adaptation Model of High-Performance Work Teams”, *Journal of Applied Psychology*, vol.104, n°3, 2019, pp.433: <https://doi.org/10.1037/apl0000372>.

- EAGLY, A.H. & S.J. KARAU, "Role Congruity Theory of Prejudice toward Female Leaders", *Psychological Review*, vol.109, n°3, 2002, pp.573-598: <https://psycnet.apa.org/doi/10.1037/0033-295X.109.3.573>.
- FOSCHI, M., "Gender and Double Standards for Competence", pp.181-207 in C.L. Ridgeway (ed.), *Gender Interactions and Inequality*, New York, Springer Science + Business Media, 1992a.
- FOSCHI, M., "Status Characteristics, Standards, and Attributions", pp.50-71 in J. Berger, M. Zelditch & B. Anderson (eds.), *Sociological Theories in Progress : New Formulations*, Thousand Oaks, CA, Sage Publications, 1992b.
- FOSCHI, M., "Double Standards for Competence : Theory and Research", *Annual Review of Sociology*, vol.26, n°1, 2000, pp.21-42 : <https://doi.org/10.1146/annurev.soc.26.1.21>.
- FOSCHI, M., L. LAI & K. SIGERSON, "Gender and Double Standards in the Assessment of Job Applicants", *Social Psychology Quarterly*, 57(4), 1994, pp.326-339 : <http://www.jstor.org/stable/10.2307/2787159>.
- FOSCHI, M., G.K. Warriner & S.D. Hart, "Standards, Expectations, and Interpersonal Influence", *Social Psychology Quarterly*, vol.48, n°2, 1985, pp.108-117 : <https://doi.org/10.2307/3033606>.
- GARCIA-PRieto, P., E. BELLARD & S.C. SCHNEIDER, "Experiencing Diversity, Conflict, and Emotions in Teams", *Applied Psychology*, vol.52, n°3, 2003, pp.413-440: <https://doi.org/10.1111/1464-0597.00142>.
- GERBER, G.L., "Status in Same-Gender and Mixed-Gender Effects on Personality Attributions", *Social Psychology Quarterly*, vol.59, n°4, 1996, pp.350-363 : <https://doi.org/10.2307/2787076>.
- GLOCK, S. & H. KLEEN, "Shifting Standards for Female Ethnic Minority Students ? Evidence from Two Experimental Studies", *Educational Research & Evaluation*, 2020, pp.1-19.
- HACKMAN, J.R., "A Real Team", pp.36-60 in J.R. Hackman (ed.), *Leading Teams : Setting the Stage for Great Performances*, Cambridge, MA, Harvard Business School Press, 2002.
- HACKMAN, J.R. & N. KATZ, "Group Behavior and Performance", pp.1208-1251 in L.G. Fiske, Susan T. Gilbert & T. Daniel (eds.), *Handbook of Social Psychology*, volume 2 (5th ed.), Hoboken, NJ, John Wiley & Sons, Inc., 2010.
- HAIR, J.F., Jr. & M. SARSTEDT, "Factors versus Composites : Guidelines for Choosing the Right Structural Equation Modeling Method", *Project Management Journal*, vol.50, n°6, 2019, pp.619-624 : <https://doi.org/10.1177/8756972819882132>.
- HASTINGS, D.W., "Drowning in Inequalities : Swimming and Social Justice", *Journal of Black Studies*, vol.36, n°6, 2006, pp.894-917 : <https://doi.org/10.1177/0021934705283903>.
- HERRING, C., "Does Diversity Pay ? Race, Gender, and the Business Case for Diversity", *American Sociological Review*, vol.74, n°2, 2009, pp.208-224 : <https://doi.org/10.1177/000312240907400203>.
- HEWSTONE, M., M. RUBIN & H. WILLIS, "Intergroup Bias", *Annual Review of Psychology*, vol.53, 2002, pp.575-604 : <https://doi.org/10.1146/annurev.psych.53.100901.135109>.
- HOCHBERG, J., *Evaluating High Performing Female Colleagues : The Roles of Race, Gender, and Task Performance* [unpublished doctoral dissertation], University of Michigan, 2020.
- HOFFMAN, L. & N. MAIER, "Quality and Acceptance of Problem Solutions by Members of Homogeneous and Heterogeneous Groups", *Journal of Abnormal & Social Psychology*, vol.62, n°2, 1961, pp.401-407 : <https://doi.org/10.1037/h0044025>.
- HORWITZ, S.K. & I.B. HORWITZ, "The Effects of Team Diversity on Team Outcomes : A Meta-Analytic Review of Team Demography", *Journal of Management*, vol.33, n°6, 2007, pp.987-1015 : <https://doi.org/10.1177/0149206307308587>.
- JACKSON, S.E., A. JOSHI & N.L. ERHARDT, "Recent Research on Team and Organizational Diversity : SWOT Analysis and Implications", *Journal of Management*, vol.29, n°6, 2003, pp.801-830 : <https://doi.org/10.1016/S0149-2063>.

JEHN, K.A., G.B. NORTHCRAFT & M.A. NEALE, “Why Differences Make a Difference : A Field Study of Diversity, Conflict and Performance in Workgroups”, *Administrative Science Quarterly*, vol.44, n°4, 1999, pp.741-763 : <https://doi.org/10.2307/2667054>.

KALKHOFF, W., D. MELAMED, J. POLLOCK, B. MILLER, J. OVERTON & M. PFEIFFER, “Cracking the Black Box : Capturing the Role of Expectation States in Status Processes”, *Social Psychology Quarterly*, vol.83, n°1, 2020, pp.26-48 : <https://doi.org/10.1177/0190272519868988>.

KOCHAN, T., K. BEZRUKOVA, R.J. ELY, S.E. JACKSON, A. JOSHI, K. JEHN & J. LEONARD, “The Effects of Diversity on Business Performance : Report of the Diversity Research Network”, *Human Resource Management*, vol.42, n°1, 2003, pp.3-21 : <https://doi.org/10.1002/hrm.10061>.

LAIRAMORE, C., D. MORRIS, R. SCHICHTL, L. GEORGE-PASCHAL, H. MARTENS, A. MARAGAKIS & A. BRUENGER, “Impact of Team Composition on Student Perceptions of Interprofessional Teamwork : A Six-Year Cohort Study”, *Journal of Interprofessional Care*, vol.32, n°2, 2018, pp.143-150 : <https://doi.org/10.1080/13561820.2017.1366895>.

LAU, D.C. & J.K. MURNIGHAN, “Demographic Diversity and Faultlines : The Compositional Dynamics of Organizational Groups”, *The Academy of Management Review*, vol.23, n°2, 1998, pp.325-340 : <https://doi.org/10.2307/259377>.

LAU, D.C. & J.K. MURNIGHAN, “Interactions within Groups and Subgroups : The Effects of Demographic Faultlines”, *Academy of Management Journal*, vol.48, n°4, 2005, pp.645-659 : <https://doi.org/10.5465/amj.2005.17843943>.

MALIK, A., H. AZIZ & T. AHMED, “Power, Intrateam Conflicts, Conflict Contagion and its Impact on Performance”, *Information Management & Business Review*, vol.11, n°2 (I), 2019, pp.21-26 : [https://doi.org/10.22610/imbr.v11i2\(I\).2879](https://doi.org/10.22610/imbr.v11i2(I).2879).

MANNIX, E. & M.A. NEALE, “What Differences Make a Difference?: The Promise and Reality of Diverse Teams in Organizations”, *Psychological Science in the Public Interest*, vol.6, n°2, 2005, pp.31-55 : <https://doi.org/10.1111/j.1529-1006.2005.00022.x>.

MATHIEU, J., M.T. MAYNARD, T. RAPP & L. GILSON, “Team Effectiveness 1997-2007 : A Review of Recent Advancements and a Glimpse into the Future”, *Journal of Management*, vol.34, n°3, 2008, pp.410-476 : <https://doi.org/10.1177/0149206308316061>.

METRICK, J. & G. COWAN, “Gender Stereotypes and Predictions of Performance : Women in Air Combat”, *Journal of Social Behavior & Personality*, vol.11, n°5, 1996, pp.105-120.

MILLIKEN, F.J. & L.L. MARTINS, “Searching for Common Threads: Understanding the Multiple Effects of Diversity in Organizational Groups”, *Academy of Management Review*, vol.21, n°2, 1996, pp.402-433 : <https://doi.org/10.5465/amr.1996.9605060217>.

MO, S., C.D. LING & X.Y. XIE, “The Curvilinear Relationship between Ethical Leadership and Team Creativity : The Moderating Role of Team Faultlines”, *Journal of Business Ethics*, vol.154, n°1, 2019, pp.229-242 : <https://doi.org/10.1007/s10551-016-3430-1>.

NEWHEISER, A.-K. & J.F. Dovidio, “Individual Differences and Intergroup Bias : Divergent Dynamics Associated with Prejudice and Stereotyping”, *Personality & Individual Differences*, vol.53, n°1, 2012, pp.70 : <http://doi:10.1016/j.paid.2012.02.024>.

OFFICE OF THE UNDER-SECRETARY OF DEFENSE FOR PERSONNEL AND READINESS, *Population Representation in the Military Services : Fiscal Year 2017 Summary Report* : <https://www.cna.org/pop-rep/2017/summary/summary.pdf>.

OPARAH, J.C., “Feminism and the (Trans)Gender Entrapment of Gender Non-Conforming Prisoners”, *UCLA Women’s Law Journal*, vol.18, n°2, 2012, pp.239-271.

- PEARSALL, M.J., A.P.J. ELLIS & J.M. EVANS, “Unlocking the Effects of Gender Faultlines on Team Creativity : Is Activation the Key ?”, *Journal of Applied Psychology*, vol.93, n°1, 2008, pp.225-234 : <https://doi.org/10.1037/0021-9010.93.1.225>.
- PENNINGTON, C.R., D. HEIM, A.R. LEVY & D.T. LARKIN, “Twenty Years of Stereotype Threat Research: A Review of Psychological Mediators”, *PLoS ONE*, vol.11, n°1, 2016 : doi.org/10.1371/journal.pone.0146487.
- PENTLAND, A., “The New Science of Building Great Teams”, *Harvard Business Review*, April 2012, pp.60-70.
- PETTIGREW, T.F. & L.R. TROPP, “A Meta-Analytic Test of Intergroup Contact Theory”, *Journal of Personality & Social Psychology*, vol.90, n°5, 2006, pp.751 : <https://psycnet.apa.org/doi/10.1037/0022-3514.90.5.751>.
- PETTIGREW, T.F. & L.R. TROPP, “How Does Intergroup Contact Reduce Prejudice ? Meta-Analytic Tests of Three Mediators”, *European Journal of Social Psychology*, vol.38, n°6, 2008, pp.922-934 : <https://doi.org/10.1002/ejsp.504>.
- PHAM, T., R. METOYER, K. BEZRUKOVA & C. SPELL, “Show Me the Cracks in Our Teams”: Visual Representations of Demographic Diversity Faultlines. IEEE InfoVis2012 Poster Abstract, 2012.
- PITTINSKY, T.L., S.A. ROSENTHAL & R.M. MONTOYA, “Measuring Positive Attitudes toward Outgroups : Development and Validation of the Allophilia Scale”, pp.41-60 in L.R. Tropp & R.K. Mallett (eds.), *Moving beyond Prejudice Reduction : Pathways to Positive Intergroup Relations* (pp. 41-60), American Psychological Association : <https://content.apa.org/doi/10.1037/12319-002>.
- POLZER, J.T., L.P. MILTON & W.B. SWANN, “Capitalizing on Diversity : Interpersonal Congruence in Small Work Groups”, *Administrative Science Quarterly*, vol.47, n°2, 2011, pp.296-324 : <https://doi.org/10.2307%2F3094807>.
- RANDOLPH, K., *How Attitudes Affect Gender as a Status Characteristic* [Doctoral dissertation, University of North Carolina at Charlotte], 2019.
- RIDGEWAY, C.L. & J. BERGER, “Expectations, Legitimation, and Dominance Behavior in Task Groups”, *American Sociological Review*, vol.51, n°5, 1986, pp.603-617 : <https://doi.org/10.2307/2095487>.
- ROBERGE, M.-É. & R. VAN DICK, “Recognizing the Benefits of Diversity: When and How Does Diversity Increase Group Performance ?”, *Human Resource Management Review*, vol.20, n°4, 2010, pp.295-308 : <https://doi.org/10.1016/j.hrmr.2009.09.002>.
- ROBERTS, S. & D.K. BROWN, “How to Manage Gender Bias from Within: Women in Leadership”, *Journal of Business Diversity*, vol.19, n°2, 2019 : [doi:10.33423/jbd.v19i2.2057](https://doi.org/10.33423/jbd.v19i2.2057)
- ROSENTHAL, H.E.S., R.J. CRISP & M.W. SUEN, “Improving Performance Expectancies in Stereotypic Domains : Task Relevance and the Reduction of Stereotype Threat”, *European Journal of Social Psychology*, vol.37, n°3, 2007, pp.586-597 : <https://doi.org/10.1002/ejsp.379>.
- ROTHBART, M. & O.P. JOHN, “Social Categorization and Behavioral Episodes : A Cognitive Analysis of the Effects of Intergroup Contact”, *Journal of Social Issues*, vol.41, n°3, 1985, pp.81-104 : <https://psycnet.apa.org/doi/10.1111/j.1540-4560.1985.tb01130.x>.
- SAILES, G., “The Myth of Black Sports Supremacy”, *Journal of Black Studies*, vol.21, n°4, 1991, pp.480-487 : <https://doi.org/10.1177%2F002193479102100407>.
- SHIH, M., T.L. PITTINSKY & N. AMBADY, “Stereotype Susceptibility : Identity Salience and Shifts in Quantitative Performance”, *Identity*, vol.10, n°1, 1999, pp.80-84 : <https://doi.org/10.1111%2F1467-9280.00111>.
- SMITH, D.G., J.E. ROSENSTEIN, M.C. NIKOLOV & D.A. CHANEY, “The Power of Language : Gender, Status, and Agency in Performance Evaluations”, *Sex Roles*, vol.80, n°3-4, 2019, pp.159-171 : <https://doi.org/10.1007/s11199-018-0923-7>.

- STEELE, C.M. & J. ARONSON, "Stereotype Threat and the Intellectual Test Performance of African Americans", *Journal of Personality & Social Psychology*, vol.69, n°5, 1995, pp.797-811 : [doi:10.1037/0022-3514.69.5.797](https://doi.org/10.1037/0022-3514.69.5.797).
- TAJFEL, H., "Social Categorization, Social Identity, and Social Comparison", pp.61-76 in H. Tajfel (ed.), *Differentiation between Social Groups*, London, Academic Press, 1978.
- TAJFEL, H. & J.C. TURNER, "The Social Identity Theory of Intergroup Behaviour", pp.7-24 in S. Worchel & W. G. Austin (eds.), *Psychology of Intergroup Relations*, Chicago, Nelson-Hall, 1986.
- TAJFEL, H., J.C. TURNER, W.G. AUSTIN & S. WORCHEL, "An Integrative Theory of Intergroup Conflict", pp.33-37 in W.G. Austin & S. Worchel (eds.), *The Social Psychology of Intergroup Relations*, Monterey, CA, Brooks/Cole, 1979.
- THATCHER, S.M.B. & P.C. PATEL, "Demographic Fault Lines: A Meta-Analysis of the Literature", *Journal of Applied Psychology*, vol.96, n°6, 2011: pp.1119-1139. <https://psycnet.apa.org/doi/10.1037/a0024167>.
- VAN DIJK, H., M.L. VAN ENGEN & D. VAN KNIPPENBERG, "Defying Conventional Wisdom: A Meta-Analytical Examination of the Differences between Demographic and Job-Related Diversity Relationships with Performance", *Organizational Behavior & Human Decision Processes*, vol.119, n°1, 2012, pp.38-53 : <https://doi.org/10.1016/j.obhdp.2012.06.003>.
- VAN KNIPPENBERG, D., C.K.W. DE DREU & A.C. HOMAN, "Work Group Diversity and Group Performance: An Integrative Model and Research Agenda", *Journal of Applied Psychology*, vol.89, n°6, 2004, pp.1008-1022 : <https://doi.org/10.1037/0021-9010.89.6.1008>.
- VAN KNIPPENBERG, D. & M.C. SCHIPPERS, "Work Group Diversity", *Annual Review of Psychology*, vol.58, 2007, pp.515-541 : <http://dx.doi.org/10.1146/annurev.psych.58.110405.085546>.
- WAGNER, D.G. & J. BERGER, "Gender and Interpersonal Task Behaviors: Status Expectation Accounts", *Sociological Perspectives*, vol.40, n°1, 2011, pp.1-32 : <https://doi.org/10.2307%2F1389491>.
- WATSON, W.E., K. KUMAR & L.K. MICHAELSEN, "Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups", *Academy of Management Journal*, vol.36, n°3, 1993, pp.590-602 : <https://doi.org/10.2307/256593>.
- WEEKS, M., "Socioeconomic Status is Not Judged Equally: Target Race Shifts Standards in Inter-personal Judgments of SES", *European Journal of Social Psychology*, vol.49, n°2, 2019, pp.286-299 : <https://doi.org/10.1002/ejsp.2392>.
- WELLS, J.E. & T.J. AICHER, "Follow the Leader: A Relational Demography, Similarity Attraction, and Social Identity Theory of Leadership Approach of a Team's Performance", *Gender Issues*, vol.30, n°1-4, 2013, pp.1-14 : [doi:10.1007/s12147-013-9112-8](https://doi.org/10.1007/s12147-013-9112-8).
- WILLIAMS, K.Y. & C.A. O'REILLY, "Demography and Diversity in Organizations: A Review of 40 Years of Research", *Research in Organizational Behavior*, vol.20, 1998, pp.77-140.
- YZERBYT, V. & S. DEMOULIN, "Intergroup Relations", pp.1024-1083 in G.L.S.T. Fiske, & D.T. Gilbert (eds.), *Handbook of Social Psychology*, volume 2 (5th ed.), Hoboken, NJ, John Wiley & Sons, 2010 : <https://doi.org/10.1002/9780470561119.socpsy002028>.
- ZANUTTO, E.L., K. BEZRUKOVA & K.A. JEHN, "Revisiting Faultline Conceptualization: Measuring Faultline Strength and Distance", *Quality & Quantity*, vol.45, n°3, 2011, pp.701-714 : <https://doi.org/10.1007/s11135-009-9299-7>.
- ZHANG, C. & C. LIU, "Research on Team Faultlines: A Literature Review", *Journal of Human Resource & Sustainability Studies*, vol.7, n°1, 2019, pp.37-54 : [doi:10.4236/jhrss.2019.71004](https://doi.org/10.4236/jhrss.2019.71004).

Appendix A

Cadet- and Team-Level Variables

Cadet-Level Variables

Cadet gender: Cadets reported their gender on the survey's demographic portion (males = 0, females = 1).

Cadet race: Cadets reported their race on the survey's demographic portion (White = 0, non-White = 1).

Cadet class : Each cadet's class (grade or year group) was provided by the United States Military Academy's Department of Military Instruction (juniors and seniors = 0, sophomores and freshmen = 1).

Cadet diversity index (DVI): From the three status characteristics (gender, race and class) each cadet was given a score ranging between a 0 and a 3, computed as a sum of the three corresponding indicator variables to reflect their individual diversity index. For example, a White male upperclassman would be given a diversity index of 0, while a non-White female lower classman would be given diversity index of 3.

Cadet shifting standards index (SSI): Before calculating the shifting standards index for each cadet, the data had to be transposed such that each row contained the received ratings and rankings for each cadet. When the data were withdrawn from the survey instrument, each row represented a cadet's ratings and rankings of their teammates – the judgments of others made *by* that cadet. To determine each cadet's own received ratings and rankings, the data had to be transposed so that each row represented the cadet's received rankings and ratings – the judgments of others made *about* that cadet. See Exhibit 1 (next page) for an example of how the data was transposed. In this example, the upper box titled “decision-making ratings from survey” represents how the data was arranged when it was pulled from the survey. The data is arranged such that each row represents a cadet as a rater, and their row contains all of the ratings that cadet made about their teammates in each of the nine competencies. In this case, Joe's row represents Joe rating Sally a 2, Fred a 5, and Tom a 6 in the leadership competency “Decision-Making”. After transposition, each row represents a cadet as a ratee, and their row contains all of their ratings received from all of their teammates in each of the nine competencies. In the lower box titled “decision-making ratings transposed”, Joe's row now represents Joe's received ratings from Sally, a 2, Fred, a 6, and Tom, a 3.

Exhibit 1: Exhibit depicting an example of how the data was transposed, to calculate the shifting standards index (SSI).

Decision-Making Ratings From Survey Instrument					
RATEES					
RATERS		Joe	Sally	Fred	Tom
	Joe		2	5	6
	Sally	2		4	5
	Fred	6	4		5
	Tom	3	5	3	

Decision-Making Ratings Transposed					
RATERS					
RATEES		Joe	Sally	Fred	Tom
	Joe		2	6	3
	Sally	2		4	5
	Fred	5	4		3
	Tom	6	5	5	

After the data were transposed, we reverse coded all the rankings so that higher numbers represented higher rankings. For example, if a cadet named “Jane” was ranked as the best decision-maker on her 12-member team, her associated value with respect to decision-making would be a “1.” After reverse coding, Jane’s value with respect to decision-making would become a “12”.

We then generated variables representing the average rating, $rate_{ij}$, and the average ranking, $rank_{ij}$, for each cadet i on team j across all competencies C : as given by :

$$rate_{ij} = \frac{1}{C} \sum_{c=1}^C rate_{cij}$$

and

$$rank_{ij} = \frac{1}{C} \sum_{c=1}^C rank_{cij}$$

where C = the number of competencies.

Then, we computed the team average rating, \overline{rate}_j , as given by $\overline{rate}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} rate_{ij}$ and ranking, \overline{rank}_j as given by $\overline{rank}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} rank_{ij}$ where n_j = number of cadets on team.

We also computed the standard deviation of the average ratings and the average rankings of cadets on each team. The standard deviation of mean cadet ratings for team j is given by $S_j^{rate} = SD(rate_{ij})$.

The standard deviation of mean cadet rankings for team j is : $S_j^{rank} = SD(rank_{ij})$

We then standardized the average rating and ranking for each cadet on each team as given by :

$$z_{ij}^{rate} = \frac{rate_{ij} - \overline{rate_j}}{s_j^{rate}} \qquad z_{ij}^{rank} = \frac{rank_{ij} - \overline{rank_j}}{s_j^{rank}}$$

The shifting standards index (SSI) for each cadet on each team was then calculated as the difference between the cadet's standardized rating and ranking. That is,

$$SSI_{ij} = z_{ij}^{rate} - z_{ij}^{rank}$$

Team-Level Variables

Team diversity index (TDVI): The mean of all of the team members' DVIs were calculated for each team.

Team performance: The team performance score awarded by the judges and officials of the Sandhurst Competition represents the score each team earned overall in the three-day competition. The team performance scores ranged on a scale from 0 to 500 points and were calculated using a formula that consisted of the time each team required to negotiate all of the events and their adjudicated performance in each task, combined with points associated with their earned penalties. (A description of possible penalties is available upon request.) Point scores in each of the events were standardized, to facilitate the comparative analysis across events.

Appendix B

Description of the Four Models

We fit two models, Model A and Model B, to answer Research Questions 1 and 2. We fit two models, Model C and Model D, to answer Research Question 3.

Model A

Cadet level :

$$SSI_{ij} = \beta_{1j}(DVI_{ij}) + \varepsilon_{ij}$$

$$\varepsilon \sim N(0, \sigma^2)$$

Team level :

$$\beta_{1j} = \alpha_1 + \zeta_{ij}$$

$$\zeta \sim N(0, \tau)$$

β_{1j} = The change in the mean of SSI for every one unit change in cadet DVI (i.e., global stereotyping) on team j .

α_1 = The average of global stereotyping across all teams corresponding to a one- unit difference in cadet DVI.

$Var(\beta_{1j})$ = The variance of global stereotyping across all teams corresponding to a one-unit difference in cadet DVI.

Model B

Cadet level :

$$SSI_{ij} = \beta_{2j}(gender_{ij}) + \beta_{3j}(race_{ij}) + \beta_{4j}(class_{ij}) + \varepsilon_{ij}$$

$$\varepsilon \sim N(0, \sigma^2)$$

Team level :

$$\beta_{2j} = \alpha_2 + \zeta_{2j}$$

$$\beta_{3j} = \alpha_3 + \zeta_{3j}$$

$$\beta_{4j} = \alpha_4 + \zeta_{4j}$$

$$\zeta \sim MVN(0, \tau)$$

β_{2j} is the level of gender stereotyping (difference in mean SSI_{ij} for females compared to males) on team j , controlling for race and class.

β_{3j} is the level of race stereotyping (difference in mean SSI_{ij} for non-Whites compared to Whites) on team j , controlling for gender and class.

β_{4j} is the level of cadet class stereotyping (difference in mean SSI_{ij} for third- and fourth-class cadets compared to first- and second-class cadets) on team j , controlling for gender and race.

α_2 = average stereotyping with respect to gender across all teams

α_3 = average stereotyping with respect to race across all teams

α_4 = average stereotyping with respect to cadet class across all teams

$Var(\zeta_{2j})$ = variance of stereotyping with respect to gender across all teams

$Var(\zeta_{3j})$ = variance of stereotyping with respect to race across all teams

$Var(\zeta_{4j})$ = variance of stereotyping with respect to cadet class across all teams

All variables at Level 1 are group-mean centred. All variables at Level 2 are grand-mean centred. There is no intercept for SSI at Level 1 because SSI is standardized within each team.

Research Question 1 : Is there evidence of team-level shifting standards within teams competing in a skills competition ?

Evidence of team-level shifting standards with respect to all sources of stereotyping is determined by whether the value of α_1 in Model A is statistically significantly different from zero. Evidence of shifting standards (i.e., stereotyping) with respect to gender, race, and cadet class is determined by whether the value of α_2 , α_3 , and α_4 in Model B are statistically significantly different from zero, respectively.

Research Question 2 : Is there variability in team-level shifting standards across the teams?

To answer Research Question 2, we compared the level of variability of shifting standards across teams with respect to global stereotyping, and the mean differences in the shifting standards index for the different status characteristics variables (gender, race, and cadet class). Respectively, this variability is denoted by the values of $(var)\zeta_{1j}$, $(var)\zeta_{2j}$, $(var)\zeta_{3j}$, and $(var)\zeta_{4j}$. These values represent the variability in the level of stereotyping across all teams with respect to a diversity index and each specific source of stereotyping (gender, race, or cadet class). If these values were statistically significantly different from zero, then this is evidence of variability across teams in shifting standards with respect to the diversity index, or gender, race, or cadet class.

Model C

$$\text{Cadet level :} \quad SSI_{ij} = \beta_{1j}(DVI_{ij}) + \varepsilon_{ij}$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$\text{Team level :} \quad \beta_{ij} = \alpha_1 + \zeta_{ij}$$

$$\zeta \sim N(0, \tau)$$

$$P_j = \omega_0 + \omega_1(\beta_{1j}) + \xi_j$$

$$\xi_j \sim N(0, \tau)$$

ω_1 = difference in mean performance (P_j) corresponding to a one-unit increase in global stereotyping (i.e., β_{1j}).

Model D

Cadet level :

$$SSI_{ij} = \beta_{2j}(gender_{ij}) + \beta_{3j}(race_{ij}) + \beta_{4j}(class_{ij}) + \varepsilon_{ij}$$

$$\varepsilon \sim N(0, \sigma^2)$$

Team level :

$$\beta_{2j} = \alpha_2 + \zeta_{2j}$$

$$\beta_{3j} = \alpha_3 + \zeta_{3j}$$

$$\beta_{4j} = \alpha_4 + \zeta_{4j}$$

$$\zeta \sim MVN(0, \tau)$$

$$P_j = \omega_0 + \omega_2(\beta_{2j}) + \omega_3(\beta_{3j}) + \omega_4(\beta_{4j}) + \xi_j$$

$$\xi_j \sim N(0, \tau)$$

ω_2 = difference in mean performance (P_j) corresponding to a one-unit increase in stereotyping with respect to gender, controlling for stereotyping with respect to race and cadet class.

ω_3 = difference in mean performance (P_j) corresponding to a one-unit increase in stereotyping with respect to race, controlling for stereotyping with respect to gender and cadet class.

ω_4 = difference in mean performance (P_j) corresponding to a one-unit increase in stereotyping with respect to cadet class, controlling for stereotyping with respect to gender and race.

Research Question 3 : Is there a relationship between team-level shifting standards and team performance in the competition?

To answer Research Question 3, we fit Model C and Model D regressing a global stereotyping indicator (DVI) on team performance P_j . To counteract the problem of inflated Type-I error rates due to multiple comparisons across such many events, we used a Bonferroni-corrected α -level of 0.005.

The association between the level of global stereotyping (team level gender, race, and cadet class stereotyping) and team performance will be denoted by the value of ω_1 in Model C. The association between the individual sources of stereotyping with respect to gender, race, and cadet class and team performance will be denoted by the values of ω_2 , ω_3 , and ω_4 in Model D. The magnitude of these values and the sign of the coefficient P_j in Model D will determine the impact and the direction of the impact that stereotyping has on team performance.