TOKEN FEMALE VOICE ENACTMENT IN TRADITIONALLY MALE-DOMINATED TEAMS: FACILITATING CONDITIONS AND CONSEQUENCES FOR PERFORMANCE

ABSTRACT

When is a token female's voice incorporated into the actions of a traditionally male-dominated team and to what ends? Drawing from the tokenism, gender stereotypes, and minority influence literatures, we advance a model that specifies the conditions that facilitate token female voice enactment and when enacting her voice enhances team performance. Using a sample of active duty military men and women, we employed live observation techniques to study voice enactment in all-male teams versus female token teams (i.e., teams with a token female member) throughout a series of complex and physically demanding tasks. Our findings revealed that a) token female voice enactment was higher when team leaders possessed more favorable beliefs about women's capabilities in the military, and b) token female voice enactment enhanced team performance in more complex tasks but harmed team performance in less complex tasks. Additionally, our supplementary analyses revealed that female token teams were more reflective before engaging in action relative to all-male teams that tended to engage in agentic, "action-first" strategies. Theoretical and practical implications for facilitating female voice enactment in traditionally male-dominated contexts are discussed.

Keywords: Gender diversity; Leadership; Minority influence; Teams; Tokenism; Voice; Voice enactment

Employee voice – defined as individuals' discretionary communication of work-related ideas, suggestions, concerns, or opinions (Morrison, 2011) – is believed to be important for promoting effective team functioning. By converting private knowledge held by individual members into public knowledge for the team to use, voice ensures teams a) have up-to-date relevant task information, b) generate multiple decision alternatives, and c) engage in adaptive coordination in the face of change (Edmondson, 2003, 2012; Edmondson, Bohmer, & Pisano, 2001). Nevertheless, there are obstacles that threaten to nullify these benefits for team performance. Voice may not always be expressed (Morrison & Milliken, 2000), recognized (Burris, Detert, & Romney, 2013), or positively evaluated (Burris, 2012). Moreover, once expressed, voice may not necessarily be acted upon (Fast, Burris, & Bartel, 2014). These issues may be particularly salient for minority members who differ from the majority on observable socio-demographic features and encounter difficulties in getting their voice enacted by the team.

Take, for example, women who are newly integrated into traditionally male-dominated settings, such as the military, top management, science, and engineering (Joshi & Roh, 2009). Anecdotal evidence suggests that when it comes to voice and voice enactment in these contexts, the token female – defined as the solo female in a group of males (Kanter, 1977) – may find her voice ignored or "drowned out" (Sandberg & Grant, 2015; Tessier, 2016). If a diverse workforce generates more performance-enhancing alternatives than homogeneous teams (De Dreu & West, 2001), there is clearly a need to understand when women's voices – especially that of a token female in an all-male team – are enacted. There is also a need to understand when enacting her voice facilitates team performance, as this may reveal the costs of failing to do so.

In the voice literature, recent work has begun to explore gender differences in whose voice is recognized and given credit (Howell, Harrison, Burris, & Detert, 2015; McClean,

Martin, Emich, & Woodruff, 2018), but this research has primarily focused on women's voice frequency rather than voice enactment – i.e., its incorporation into real-time decisions related to team strategies and processes – and its potential impact on team outcomes. In the diversity literature, traditional perspectives suggest token females are evaluated negatively by majorities due to a host of perceptual, attributional, and stereotype-driven processes (Kanter, 1977; Heilman, 2012). However, other research suggests that the presence of a demographically dissimilar other can increase a team's openness to different perspectives (Phillips, Liljenquist, & Neale, 2009; Phillips, Northcraft, & Neale, 2006). These competing rationales appear to present mutually exclusive arguments for predicting token female voice enactment – and neither perspective describes the factors that may tip forces in favor of her voice enactment nor the conditions under which enacting her voice benefits team performance.

To address these limitations, our theoretical model draws on the tokenism, gender stereotype, and minority influence literatures to answer two related research questions. First, what conditions facilitate token female voice enactment in traditionally male-dominated team contexts? Second, what are the implications of enacting token female voice for team performance? As token females are often "new minorities" in otherwise all-male settings, team members' may not have prior experiences to guide interactions with the token female, and her voice is likely to be evaluated on the basis of social-group expectations informed by stereotypes. Additionally, because token females not only look different but are also likely to present viewpoints that differ from those of an all-male majority (Eagly & Wood, 1991; Feingold, 1994; Halpern, 1986), we draw on the minority influence literature to understand when a demographically dissimilar individual can shape the thoughts and actions of the majority (Moscovici, 1980; Nemeth, 1986; Phillips et al., 2006; Thomas-Hunt, Ogden, & Neale, 2003).

By integrating and extrapolating from these perspectives, we further identify leader's gender beliefs as a facilitator of token female voice enactment and task complexity as a boundary condition of when enacting her voice will facilitate team performance in female token teams.

Our work offers several contributions. We extend voice research by going beyond investigations of voice frequency (e.g., Morrison, 2014) to focus on voice enactment, a construct alluded to as an indicator of managerial receptivity (Fast et al., 2014) but not yet formally studied in the literature. By examining the conditions that facilitate token female voice enactment and the performance implications of acting on her voice, we offer new insight on when the voice of demographic minorities can impact and shape performance in majority-dominated team contexts (Howell et al., 2015). We also extend the team diversity literature by identifying boundary conditions, such as the leader's gender beliefs, that reconcile competing perspectives around whether and when token females' voices are enacted. More broadly, our focus on voice enactment investigates the micro-dynamics that enable specific individuals to affect the performance outcomes in diverse teams (van Knippenberg & Mell, 2016; van Dijk, Meyer, van Engen, & Loyd, 2016) and sheds light on the processes that underlie contingency factors of diversity effects (e.g., Plaut, 2010; van Knippenberg & Schippers, 2007). Lastly, we extend the minority influence literature, which has traditionally focused on how those expressing minority opinions produce conversion among the majority (Moscovici, 1980) but rarely examined when the opinions of newly integrated, socio-demographic tokens are enacted and shape performance.

Beyond these theoretical extensions, our work presents an important contribution because we test our model in a branch of the U.S. military that, at the time of data collection, was undergoing legislative changes to open certain all-male job categories to women for the first time in history. Women were a new and unprecedented minority in this context, and the male majority

had limited previous experience working with them on interdependent tasks (Joshi & Roh, 2009). Thus, our work offers evidence-guided and timely implications for managing newly integrated teams in a setting of national importance.

THEORETICAL BACKGROUND

Voice and Voice Enactment in Teams

In team settings, members exhibit voice when they speak up with improvement-oriented suggestions, concerns, and opinions related to the team's task (Farh & Chen, 2018; Li, Liao, Tangirala, & Firth, 2017; Morrison, Wheeler-Smith, & Kamdar, 2011). Unlike in supervisor-subordinate dyads, voice in teams is often directed towards the team leader *as well as* fellow team members because teams are characterized by interdependence, diffusion of expertise, and distributed authority for action (Li et al., 2017; Morrison et al., 2011; Morrison, 2014). Moreover, because team interactions revolve around the accomplishment of dynamic team tasks (Kozlowski & Ilgen, 2006; Marks, Mathieu, & Zaccaro, 2001), the status quo upon which voice attempts to improve reflects the team's current reasoning or approach to the team task (Premeaux & Bedeian, 2003). Thus, in theory, voice helps the team iteratively improve its current approach to accomplishing its tasks (Edmondson, 2003; Edmondson et al., 2001; Farh & Chen, 2018).

Voice is considered an essential source of health and functioning for teams (Mesmer-Magnus & DeChurch, 2009; Nemeth, Connell, Rogers, & Brown, 2001), and when members fail to exercise voice, there can be serious negative consequences for team effectiveness and performance (Argyris, 1977; Janis, 1972; Perlow & Williams, 2003). Yet, scholars are increasingly realizing that simply engaging in voice, does not guarantee that one's suggestions will impact the team in consequential ways. For instance, ignoring or contradicting a member's voice limits its impact. Likewise, simply valuing voice does not imply action. On the other hand,

voice enactment – which we conceptualize as the extent voice is incorporated into the team's approach to the task – has the potential to directly shape team outcomes.

Because voice enactment requires voice recipients to take action on the speaker's suggestion, it reflects a type of interpersonal influence (Correll & Ridgeway, 2006). Still, voice enactment can be distinguished from other types of influence that reflect changes in opinion, thought, or behavior that may or may not have occurred in response to voice, such as conformity to social norms or obedience to authority (Cialdini & Goldstein, 2004). Voice enactment is also distinct from informal leadership, which occurs when a member who is not the formal leader acts to satisfy team needs and enhance team effectiveness (Morgeson, DeRue, & Karam, 2010). Though a team member's voice is a type of change-oriented behavior (DeRue, Nahrgang, Wellman, & Humphrey, 2011), informal leadership also encompasses non-voice forms of influence, such as role modeling or building a positive socioemotional team climate (Lord, 1977). Thus, we prefer the label "voice enactment" because it more precisely captures the change-oriented manner in which a member influences the team toward team goals.

Given our focus on voice enactment and its implications for team performance, we make two assumptions going forward. First, although voice can encompass suggestions, concerns, and other opinions (Liang, Farh, & Farh, 2012; Maynes & Podsakoff, 2014), we focus on suggestion-based forms of voice, which convey proposed actions and alternatives (LePine & Van Dyne, 1998; Premeaux & Bedeian, 2003) that – if enacted by the team – can have direct and immediate implications for the team's task accomplishment (Hollenbeck, Ilgen, Sego, Hedlund, Marjor, & Phillips, 1995). Opinions or concerns may play an important role in shaping team processes (Liang, Shu, & Farh, 2018), but they do not always present actionable ideas. Second, we acknowledge that voice enactment requires the speaker to first express voice. That is, voice not

expressed cannot be enacted. Thus, while potential gender differences in voice frequency (Sandberg & Grant, 2015) may have downstream consequences for voice enactment, the focus of our work is to investigate what happens to token females' voice *after* it is expressed.

Token Female Voice Enactment in Male-Dominated Teams

To understand how a homogeneous male majority may react to token female voice, we integrate perspectives from the tokenism, gender stereotypes, and minority influence literatures. Tokens refer to individuals who are identified by an ascribed characteristic that is shared by few others in the group and carries a set of assumptions about culture, status, and behavior (Kanter, 1977). Because tokens are lone representatives of their ascribed category, their perceived differences from the dominant group tend to be exaggerated by majority members, and their actions are often scrutinized through the lens of stereotypes (Rustad, 1982; Yoder, Adams, & Prince, 1983). This is likely in the case of token females, whose lower status in society's sex hierarchy reduces their power to defy a male majority's assertions of gender stereotypes (Fairhurst & Snavely, 1983). Likewise, in contexts where token females are "newly" integrated into male-dominated settings, interpersonal knowledge and experience has not yet been established. Thus, gender stereotypes likely guide the male majority's initial perceptions of and interactions with the token female (Finseraas, Johnsen, Kotsadam, & Torsvik, 2016).

The gender stereotype literature indicates how a number of cognitive and attributional forces may work against token female voice enactment. First, descriptive and prescriptive gender stereotypes imply that women are and should be communal, conforming, nurturing, and self-effacing (Heilman, 2012). To the extent the male majority expects token females to conform to these pre-existing generalizations about their category (Kanter, 1977), they are likely to perceive her engagement in an agentic, change-oriented behavior such as voice as incongruent with their

expectations (Eagly, 1987; McClean et al., 2018). Thus, women may incur penalties for voicing their suggestions or have those suggestions negatively evaluated by the male majority (Heilman, 2001; Rudman & Glick, 2001). Second, the hierarchical dimension of gender stereotypes produces expectations that women are less competent than men (Ridgeway & Correll, 2004), which in turn may lead male counterparts to ignore or dismiss her suggestions, deeming them as unlikely to benefit the team. Finally, because being a token exaggerates perceived differences between the dominant male majority and the token female, efforts may be made to socially isolate her as an outsider, further reducing the likelihood of trusting and enacting her voice (Kanter, 1977; Yoder et al., 1983). For these reasons, it is plausible that token female voice enactment will be low in traditionally male-dominated teams.

In contrast, the literature on minority influence suggests that perceived dissimilarity from a minority individual can facilitate majority openness to the minority's perspectives. Specifically, research on minority influence (e.g., De Dreu & De Vries, 2001; Moscovici, Mucchi-Faina, & Maass, 1994) argues that when an individual presents a view that is perceived as different from the rest of the group, members of the majority are *more* likely to process the content of his or her view in a systematic way. That is, although the majority may be motivated to reject the minority position (Asch, 1956; Deutsch & Gerard, 1955; Nemeth, 1986), the act of resisting it leads majority individuals to carefully evaluate the arguments raised. In turn, this systematic processing illuminates the true merits of the message, produces a deeper understanding of the issue, and ultimately increases the influence of the minority (Forgas & Williams, 2001; Maass, West, & Cialdini, 1987; Mugny, 1982; Mugny, Butera, Sanchez-Mazas, & Perez, 1995; Nemeth & Wachtler, 1983). Moreover, there is research suggesting that attitudinal changes due to minority influence tend to be accompanied by changes in behavioral

intentions (Martin, Martin, Smith, & Hewstone, 2007), suggesting an increased likelihood of enacting a minority member's voice.

Although most of the above work was conducted in relation to minority opinion holders, some research suggests that the basic mechanism of increased systematic processing generalizes to individuals known as "double minorities" (i.e., those who also differ from the majority in terms of their social category; Moscovici, 1985). For example, Phillips (2003) argued that a minority opinion of a social category outgroup member may engender greater information processing in the team compared to if that opinion came from an ingroup member (Crano & Avaro, 1998; Crano & Hannula-Bral, 1994; Gorenflo & Crano, 1989). This is because majority members expect other ingroup members to hold similar opinions to their own. As such, a deviant opinion from an ingroup member produces a certain level of dissonance or relational discomfort and is met with negative evaluation. In contrast, because majority members are less concerned about their social relationship with an outgroup member and do not expect categorically dissimilar others to share the majority's perspectives and opinions, exposure to a minority individual's dissenting opinion generates less discomfort that may otherwise interfere with processing the minority perspective (Phillips, 2003). In fact, the presence of categorically dissimilar others may heighten the group's awareness that demographic minority members have different information or ideas that may potentially benefit the team's task accomplishment, which in turn may legitimize unique perspectives voiced by a minority member (Phillips, 2003).

Indirectly supporting these claims, research has shown that when teams contained a demographically dissimilar individual, the team was more likely to make use of informational diversity (Phillips, 2003; Phillips, Mannix, Neale, & Gruenfeld, 2004). Likewise, Thomas-Hunt et al. (2003) found that socially connected ingroup members gave greater attention to the unique

information contributed by socially isolated outgroup members than to information contributed by other ingroup members. Research has also shown that groups with greater surface-level diversity (e.g., age, gender, race) exhibited more task engagement, shared more information, and expressed more dissenting opinions compared to homogeneous groups (Loyd, Wang, Phillips, & Lount, 2013; Phillips & Loyd, 2006; Phillips et al., 2006). Although none of this work looked specifically at how these effects translate into token female voice enactment in an otherwise all-male team, they nonetheless provide a counter-narrative in which a token female is likely to have her voice enacted, precisely because of her token status and minority perspective.

Taken together, contrasting forces appear to shape the male majority's likelihood of enacting a token female's voice. On the one hand, the majority's perceptions and attributions of a stereotyped token female may produce negative evaluations that render her voice enactment less likely. On the other hand, expectations of unique information guided by minority influence mechanisms may cause the male majority to engage in greater information processing and consideration of a token female's suggestions, making her voice enactment more likely. To reflect these divergent predictions in the literature, we offer the following competing hypotheses:

Hypothesis 1a: Token female voice enactment will be lower than a comparison male.

Hypothesis 1b: Token female voice enactment will be higher than a comparison male.

Leader's Gender Beliefs as a Facilitator of Token Female Voice Enactment

We now turn to consider factors that likely turn the tide in favor of token female voice enactment. It seems that if factors exist to counteract negative perceptions, attributions, and stereotypes about female members and at the same time legitimize the value of systematically processing her view points, token female voice enactment is more likely to occur (Ely & Thomas, 2001). One such factor may be favorable gender beliefs. In particular, beliefs that

women are competent and capable in the relevant work domain or occupation are most pertinent for facilitating token female voice enactment because they counteract the status-based gender stereotype that women are less competent than men (Ridgeway & Correll, 2004). Indeed, research suggests that when perceivers possess counter-stereotype expectations about an actor's social group, normative fit is reduced, and they are less likely to categorize the actor based on his or her social categories (van Knippenberg, De Dreu, & Homan, 2004). At the same time, the minority influence literature suggests that when the majority holds positive attributions of the minority member, majority members are more likely to systematically process the message, enabling minority influence to occur (Clark & Maass, 1988; Moskowitz, 1996). In this way, beliefs that women are competent and capable in the relevant work domain also facilitate minority influence mechanisms because such perceptions are foundational for increased advice taking and interpersonal influence (Mayer, Davis, & Schoorman, 1995; Nebus, 2006).

Moreover, we propose that the beliefs of the *team leader* are most essential because of their formal position in the hierarchy and their control over whose voice is enacted in the course of task accomplishment (Hollenbeck et al., 1995; Zaccaro, Rittman, & Marks, 2001). The highly visible nature of their attitudes and actions also allows team leaders to disproportionately influence the way the token female's voice is perceived in the team. We believe leaders' positive gender beliefs may facilitate token female voice enactment in three ways. First, to the extent in which leaders holding these beliefs exhibit more positive affective and evaluative reactions toward a token female's suggestions (van Knippenberg et al., 2004), leaders offer reassurance to the rest of the group that the token female is competent and attending to her voice will result in valuable gains for enhancing task performance. Indeed, research has shown that a high-status individual's endorsement can help legitimize the contributions of individuals who are relatively

unknown or foreign entities (Lin, 1999). This may reduce the male majority's tendency to negatively stereotype the token female and marginalize her voice (Westphal & Milton, 2000) or assumed lower competency (Ridgeway & Correll, 2004).

Second, leaders who react favorably to the token female's voice signal to the male majority that she – like them – is a contributing member of the team. That is, members who see their leaders responding in more positive and inclusive ways toward the token female may be less likely to categorize her as an outgroup member and instead see her as a fellow team member who shares the common objective of performing well on the team task. Indeed, research suggests that when similarities are highlighted between minority and majority group members, minority individuals can avoid or minimize outgroup biases and be evaluated more positively (Moscovici, 1985; Volpato, Maass, Mucchi-Faina, & Vitti, 1990).

Finally, leaders who ascribe to more positive gender beliefs can facilitate token female voice enactment by protecting her and the rest of the group from social pressures to conform to the majority. Indeed, majority members may find it interpersonally risky to adopt the minority position (Moscovici, 1980), as doing so publicly presents a break from conformity with the majority. The presence of a leader who not only reduces this social pressure but also encourages the group to consider opposing perspectives is critical in ensuring that the minority is heard and respected (Maier & Solem, 1952; Mannix & Neale, 2005).

In contrast, leaders who hold less favorable beliefs about women and their competence and capabilities in a given domain are not only likely to legitimize and strengthen negative evaluations that reduce the likelihood of token female voice enactment in male-dominated teams, but also minimize the team's willingness to systematically process and be influenced by her minority perspective. Thus, we expect that:

Hypothesis 2: Leaders' gender beliefs moderate the effect of gender on voice enactment, such that female tokens are likely to have higher voice enactment than a comparison male when leaders' gender beliefs are high than when they are low.

Task Complexity as a Boundary Condition of When Token Female Voice Enactment Enhances Team Performance

Assuming sufficient levels of token female voice enactment, there are several reasons why her voice may have differential implications for team performance if enacted. First, the token female's voice likely differs in content from that of the male majority and thus contributes to the informational diversity in the team. Scholars have long argued that due to social role differences in prior experiences, socialization, and societal expectations (Eagly & Wood, 1991; Feingold, 1994), men and women often come to develop different skills, attitudes, and approaches to task accomplishment. Men, for example, tend to be more agentic, focusing on execution and attacking problems quickly and assertively (Eagly, 1987), whereas women tend to be more interpersonally facilitative, collaborative, and democratic in their approach to problem-solving (Eagly & Johannesen-Schmidt, 2001). To the extent these differences in how men and women approach and process tasks also result in differing content of suggestions, we can expect that, compared to enacting voice of the male majority, enacting token female voice will have differential implications for team performance.

Second, there may be differences in the way token female voice is cognitively processed. Research suggests that the perceived dissimilarity of a minority opinion can stimulate greater systematic processing of the opinion and produce deeper understanding of the issue (De Dreu & De Vries, 2001; Moscovici et al., 1994). This is even more likely when minority opinions are expressed by categorically dissimilar individuals because the majority *expects* these individuals

to possess different information and ideas (Phillips, 2003; Phillips & Loyd, 2006). Thus, we expect that this systematic processing and enactment of the token female's voice may produce a fundamentally different understanding of the task that breaks uniformity, undermines the dominant model, and introduces uncertainty by giving complexity to the situation (Moscovici & Mugny, 1983). Indeed, work conducted by Nemeth and colleagues found that compared to groups exposed to majority views, those exposed to minority views utilized more strategies for problem solving (Nemeth & Kwan, 1985), engaged in broader information search (Nemeth & Rogers, 1996), exhibited more originality (Nemeth & Kwan, 1987), and detected solutions that otherwise would have gone undetected (Nemeth, 1986; Nemeth & Wachtler, 1983).

These differential processes help explain when enacting the token female's voice may facilitate team performance. Research has long theorized that the effects of divergent or convergent processes on the performance of diverse teams depends greatly on the task at hand, such that "group performance should benefit primarily from diversity on complex and non-routine information-processing and decision-making tasks…" (van Knippenberg et al., 2004: 1012). Specifically, because complex tasks require diagnosing non-routine problems and generating unique and innovative solutions to those problems (De Dreu & Weingart, 2003; Morgeson & Humphrey, 2006), we expect token female voice enactment will facilitate team performance. This is not because her voice is inherently better-suited or superior for addressing complex tasks, but because enacting her voice contributes to divergent thinking in the team and deeper information processing relative to proceeding with an unchallenged majority opinion.

In contrast, low complexity tasks describe routine tasks that typically have an easily identified, effective operating procedure (De Dreu & Weingart, 2003). Because such tasks do not require teams to engage in deep and deliberate processing of task information nor generate

innovative solutions to be successful, the informational benefits derived from enacting and systematically processing divergent viewpoints may be neutralized. Moreover, because the elaboration and integration of diverse perspectives requires time and consumes resources, the costs of such processes may cause otherwise efficient work procedures to be unnecessarily delayed or abandoned in favor of more complex ones (Schwenk, 1990). Thus, in low complexity tasks where time is a fixed resource, token female voice enactment may harm performance.

In sum, we predict that the relationship between gender, voice enactment, and team performance is contingent upon task complexity. Specifically, we expect that enacting token female voice will positively impact team performance when task complexity is high, but negatively impact team performance when task complexity is low. We do not expect this same pattern of results for enacting the voice of a comparison male.

Hypothesis 3: The relationship between token female voice enactment and team performance is contingent upon task complexity such that this relationship is positive when task complexity is high and negative when task complexity is low.

METHODS

Sample and Procedures

Our sample consisted of 173 enlisted personnel (private to corporal, with 0-6 years of service) in the United States Marine Corps. To achieve a causal understanding of how token women newly integrated into an all-male team might fare with respect to voice enactment, male participants were recruited from all-male direct-combat specialties within this branch of the military, and female participants were recruited from combat support specialties that were already gender integrated, such as administration, logistics, and motor transport. Participants were recruited via flyers for a team decision-making study and offered gift cards totaling $20

USD as compensation. There was no indication from the recruitment materials or procedure that the study was about gender or voice. In our final sample, 8.2% of our participants were female, which is comparable to the statistic of 6.8% active duty women serving in this branch of the military (Statistic Brain, 2017). Participants' average age was 21.3 years, 61.2% self-identified as White, 17.7% as Hispanic, 8.8% as Black, and 12.3% indicated "Other" as their ethnic group.

We conducted our study in a controlled training facility on a military base. Prior to arriving at the research site, participants completed an online survey that measured demographics and beliefs about women in the military. Upon arrival, participants were randomly assigned to 39 all-male or female token teams of four or five members, which reflected the standard size of combat teams in this branch of the military. Twenty-two teams (56.4%) were composed of four members, and 12 (30.1%) contained a token female. Participants were dressed in identical uniforms, asked to remove all rank insignia prior to their arrival to de-emphasize existing hierarchies, and randomly assigned a color-marked helmet for the purposes of identification – blue, green, red, or yellow, and black in the case of five-person teams. An exception to this was if a team contained a token female, who was always assigned the yellow helmet.

Each team attempted four tasks of varying task complexity, producing a within-subjects design that generated 156 team observations in which each team served as its own control comparison across the different tasks. Forty-eight observations involved female token teams, and 108 observations involved all-male teams. The four tasks were adapted from standardized tasks used to train and evaluate officer candidacy in this branch of the military. Because the content of the tasks themselves were high fidelity simulations of what participants would likely encounter when deployed in combat, all tasks contained a physical component. The order of tasks attempted by each team was counter-balanced and randomly assigned. Additionally, to mimic the

hierarchical structure of combat units, each member was randomly appointed as the "formal leader" of one of the four tasks (in five-person teams, one team member did not have the opportunity to lead).

Description of Tasks

The current consensus in the literature is that tasks high in complexity do not have obvious solutions and therefore require innovative approaches (De Dreu & Weingart, 2003; Morgeson & Humphrey, 2006). Thus, in our study, a task was considered high on *task complexity* if the solution to the scenario was not readily apparent. Tasks low on task complexity had relatively straightforward solutions. We describe each of our four tasks in detail below.

"Casualty Rescue." (Appendix Task A.) The objective of this task was to cross a "booby-trapped" room, climb through a 32-inch wide hole in the wall three feet off the ground, locate and strap a 90-pound dummy to a stretcher, and move the dummy and stretcher back through the hole to the starting point – all while avoiding off-limit areas. This task was low on complexity because the solution – to use the provided 8-foot long wooden plank as a bridge to safely cross into the first room, an up-ramp to access the hole in the wall, and a down-ramp from the hole in the wall into the room with the dummy – was relatively straightforward.

"Building Resupply." (Appendix Task B.) The objective of this task was to transport five 17-pound ammunition containers and all members of the team into a building through a window that was eight feet off the ground – all while avoiding off-limit areas. The only way into the building was to scale a set of parallel benches and a stone pedestal measuring approximately six feet in height. Equipped with a 10-foot long plank, successful teams passed the plank back and forth between the two segments of the task until all members of the team and the five

ammunition containers were in the building. This task was low in complexity because the solution was straightforward.

"Medicine Delivery." (Appendix Task C.) The objective of this task was to transfer a 5-pound container from one side of a large room to the other side without throwing it or stepping into a large, off-limit space separating the two sides. Teams were split such that half the team started on one side of the room with the container and the other half was on the opposite side to receive the container. A variety of materials were scattered in adjacent rooms, including some distractor items, such that teams had to search and identify relevant and usable items. Key resources were a 40- and 18-foot long rope, a 6-foot long plank, and a hook. The most common solution implemented by teams was to combine the two lengths of rope to devise a pulley system or a zip line via the ceiling rafters to gently transfer the container across the divide. This task was high in complexity because the solution was not readily obvious, and the presence of distractors required the team to deliberate on which resources were essential and how to fashion those resources into a workable solution.

"Chemical Walls." (Appendix Task D.) The objective of this task was to transfer five 40-pound water jugs and all team members over a 10-foot high wall with two perpendicular sides. All surfaces of the wall and areas of the ground near the wall were marked as off-limits. Only the top of the wall was safe to touch. Teams were given a variety of materials to use, including some distractor items, requiring teams to identify relevant and usable items. Key resources were a 6- and 8-foot long plank. The most common solution implemented by teams was to lay both planks on top of the L-shaped corner of the wall to create a platform onto which members could safely lift themselves and the water jugs. This task was high in task complexity because the solution

was not obvious, and the presence of distractors required the team to deliberate on which

resources to use and how they could produce a workable solution.

Manipulation checks for task complexity. We consulted 17 doctoral students in a

research-intensive management department as subject matter experts (SMEs). Given a definition

of task complexity and details about each task (including the initial set up, materials provided,

and most common solution), SMEs rated each task on three items adapted from Morgeson and

Humphrey (2006) about the extent the task required "generating unique ideas," "generating

innovative solutions," and "cognitive processing" ($\alpha = .91$). The SMEs rated the two high

complexity tasks higher than the two low complexity tasks ($M_{\text{Casualty Rescue}} = 2.90$, $M_{\text{Building Resupply}}$

$= 3.14$, $M_{\text{Medicine Delivery}} = 4.10$, and $M_{\text{Chemical Walls}} = 3.84$). Results also showed that the average

SME ratings of the two high complexity tasks ($M = 3.97$, $SD = .94$) differed significantly from

the SME ratings of the two low complexity tasks ($M = 3.02$, $SD = 1.16$; $F_{(1, 66)} = 13.78$, $p < .05$).

Moreover, the SMEs did not differentiate between the two low complexity tasks nor the two high

complexity tasks. Thus, we coded each task as high or low in complexity accordingly.[1]

Measures

   Task Complexity. We coded each tasks as $0 =$ low complexity, or $1 =$ high complexity.

   Gender Composition. We coded each team as $0 =$ an all-male team, or $1 =$ a female

token team (i.e., one lone female with all-male teammates).

   Voice and Voice Enactment. We assessed voice and voice enactment using observer

ratings from four hypothesis-blind doctoral students. Using the color-marked helmets to identify

participants, observers coded in real time who voiced a suggestion and whether the team acted on

---

[1] We also analyzed our data by substituting the binary coding to task complexity with the SME's average scores across the three task complexity items ($M_{\text{Casualty Rescue}} = 2.90$, $M_{\text{Building Resupply}} = 3.14$, $M_{\text{Medicine Delivery}} = 4.10$, and $M_{\text{Chemical Walls}} = 3.84$). Results were consistent with that of our original analyses.

it. Consistent with voice being defined as suggested actions intended to improve the team's task accomplishment (Premeaux & Bedeian, 2003), observers coded voice as instances where members proposed a new plan for accomplishing an element of the task. Suggestions were strategic (e.g., "How about we try to get the rope to the other side and just sort of 'zip-line' it?") or tactical (e.g., "Lift your end onto the plank and we'll see-saw it"). Examples of non-suggestions included coordinating or directing behaviors that requested specific actions from specific members, or assigning members to roles and subtasks (Bales, 1950; Lord, 1977).

Each task was assigned a primary observer who coded all voice and voice enactment that occurred in that task. To establish the validity of the primary observer's ratings, we assessed inter-rater agreement and reliability for voice enactment using 201 person episodes[2] (29% of our total observations) in which a second observer simultaneously coded research participant behavior. Because voice enactment was a continuous count variable, we calculated intra-class correlations to assess whether the ratings of primary observers were reliable. Used in this way, higher ICC values result when the correlation between two sets of ratings is high and the variance between observer ratings across targets is low (i.e., higher inter-rater agreement and reliability; Shrout & Fleiss, 1979). Results from a one-way ANOVA confirmed that the degree of agreement across observers was acceptable ($ICC_1 = .43$ and $ICC_2 = .60$ ($F_{(200,201)} = 2.50$, $p < .05$; Bliese, 2000). Thus, only the primary observers' ratings were used for data analyses.

To ensure that voiced suggestions represented a change to the status quo, we only analyzed suggestions made after the very first suggested action as voice. We did this because research has shown that the first suggestion often anchors the team's discussion and approach to

---

[2] By person episode, we mean how much each individual's voice was enacted in a given task. With 173 participants engaged in 4 tasks, the total number of person episodes observed was 173 x 4 = 692. Of these person episode observations, 201 (i.e., 29%) were double-coded and thus used to assess inter-rater agreement and reliability.

the task (Berg, 2014) and represents the initial status quo that subsequent voice aims to change and improve upon (Premeaux & Bedeian, 2003). Thus, voice frequency reflected the total number of times a member spoke up with a suggestion after the first suggestion was made, and voice enactment reflected the total number of these suggestions that were enacted. In four-person teams, we accounted for the missing values for a hypothetical fifth member's voice in our analyses by mean replacement (i.e., averaged the voice of blue, green, and red helmet members).

Leader Gender Beliefs. We assessed participants' beliefs about the capabilities of women in the military in the pre-survey prior to team composition or leader assignment with three representative items from a pool of competence-related items outlined in Boldry, Wood, and Kashy's (2001) study of gender stereotypes in a military context. These items focused on three attributes – physical fitness, leadership, and readiness to engage in military operations – that reflected the biggest objections to women in the military (Enloe, 1983; Holland, 2006). We expected that to the extent to which participants held beliefs contrary to these objections, they would hold more positive expectations of the contributions of the token female and be more likely to enact her voice. Items were measured on a 7-point scale (*1 = strongly disagree* to *7 = strongly agree*) using the stem "In your opinion, to what extent are women in the military ..." followed by three items "physically strong", "combat ready", and "capable leaders" ($\alpha = .90$).

Team Performance. The primary observer for the task evaluated each team's performance in terms of completion time, which was measured via stopwatch and recorded as soon as the team completed the task. Although speed (efficiency) and accuracy (effectiveness) are often depicted as tradeoffs (Fitts, 1954; Woodworth, 1899), completion time in our study captured both elements of performance. Across all tasks, the only way to achieve a faster completion time was to perform the task more accurately by enacting successful strategies that

advanced the task in measurable ways. In contrast, pursuing low accuracy strategies or safety

violations that incurred a 60-second time penalty resulted in a longer completion time.

Consistent with how these tasks are typically administered, each team was allocated 18

minutes per task. Teams that did not complete the task within the 18 minutes were assigned the

maximum time of 18, indicating the lowest performance. In a post-hoc examination of cases

where teams did not complete the task, we found no case where, if given a reasonable amount of

additional time, the team would have completed the task. There were also no teams that

completed the task in exactly 18 minutes. In order to create a performance index where higher

scores reflected better performance, we subtracted the team's completion time from 18. Thus,

our performance index ranged from 0 (i.e., those who did not complete the task) to 13.32 (i.e.,

those who completed the task in 4.68 minutes, the fastest time recorded).

Control Variables. We included several control variables that potentially could impact

interpersonal processes in the team and team performance aside from gender composition. Using

demographic data captured prior to team life, we controlled for diversity in rank and ethnicity

using Blau's index of heterogeneity (1977) and age diversity using the standard deviation

(Harrison & Klein, 2007). We controlled for team size, as larger teams might perform better.

Because of fatigue effects and the potential effects of accrued experience on team interactions

(Finseraas et al., 2016), we controlled for task order, which indicated whether the episode in

question was the first, second, third, or fourth task attempted. We also controlled for the physical

demand of the task. Following the Dictionary of Occupational Titles' (1977) definition of

physical demands as the estimation of the overall strength requirement of the task, we divided

the total weight of objects and human bodies to be moved from point A to point B by the number

of team members and used the average weight per team member to categorize each task as low

($\leq$ 20 lbs. of force) or high ($\geq$ 100 lbs. of force) physical demand. Consequently, we coded

Casualty Rescue and Medicine Delivery as 0 (low physicality) and Building Resupply and

Chemical Walls as 1 (high physicality). Finally, we controlled for voice frequency to isolate

variance explained in voice enactment beyond the amount of voice that was expressed.[3]

Data Analyses. Given the multilevel nature of our data (task episodes nested within

teams) and evidence that team membership accounted for significant variance in team

performance ($ICC_1 = .11$, $F_{(38, 116)} = 1.50$, $p < .05$), we used hierarchical linear modeling for our

analyses (HLM 6.2; Raudenbush, Bryk, Cheong, Congdon, & du Tiot, 2004). Each task episode

and its associated amount of voice enactment, task complexity, and team performance were

treated as Level 1 variables. Because the person in the leader position was rotated for each task

episode, leader's gender beliefs and other episodic control variables were also treated at Level 1.

Gender composition of the team and team size were treated as Level 2 variables.

To isolate the pure effects of the Level 1 variables of interest on outcome variables within

a task episode, we group mean centered task complexity, voice frequency, and voice enactment

(Hofmann, & Gavin, 1998; Enders & Tofighi, 2007). This allowed Level 1 variables to be

expressed relative to the other cases within the same level of analysis, which in our case, is the

task episode. For example, voice enactment is considered with respect to that of other team

members' voice enactment in the same episode, yielding a slope coefficient that is unaffected by

between group differences other than the Level 2 variables under study. Group mean centering

Level 1 variables is also recommended when cross-level interactions are of interest, which was

the case as we sought to compare Level 1 relationships in all-male versus female token teams

(Bauer & Curan, 2005). Following the recommendations of Hofmann and Gavin (1998), we

---

[3] For all conclusions relevant to hypothesis testing, the omission of these controls did not change results.

reintroduced the mean level of voice enactment at Level 2 to ensure that our cross-level interactions were not spurious. To assess model fit, we used a deviance index reported in HLM analysis (Bryk & Raudenbush, 1992), where smaller deviance values indicate a better model fit. These indices are reported in Tables 4 and 5.

Analyses Involving a Comparison Male. Because token females only resided in female token teams, we used the voice enactment of a randomly assigned male counterpart from an all-male team as a comparison score as opposed to a randomly selected male from a female token team. Given that simply being exposed to minority perspectives can increase divergent thought and processing in majority members (Nemeth, 1986; Phillips et al., 2004; Nemeth & Nemeth-Brown, 2003), it is possible that input from the token female may fundamentally change the content of how majority male members think and what they choose to voice. These spillover effects, in turn, may potentially produce greater receptivity to token female voice or, in certain tasks, increase male voice enactment at the expense of enacting her voice. In either case, a comparison of token female voice enactment to male voice enactment cannot be cleanly observed when both speakers are on the same team.

Thus, to create a cleaner comparison, in Hypotheses 1 and 2, we compared the voice enactment of the token woman – who always wore a yellow helmet – to that of the male who was randomly assigned the yellow helmet in an all-male team. Moreover, when examining the effects of token female voice enactment on team outcomes (Hypothesis 3), we examined the cross-level interaction between gender composition, yellow helmet voice enactment, and task complexity on team performance, while controlling for the voice frequency and enactment of all other helmet colors (blue, green, red, and black) as well as their respective interactions with team

gender composition and task complexity. This allowed us to assess the unique contribution of the token female's voice enactment on team performance, beyond the effects of others' voices.

RESULTS

Means, standard deviations, and correlations among study variables are shown in Table 1. Average voice frequency ($F_{(1, 154)} = .43$, *n.s.*) and voice enactment ($F_{(1, 154)} = .00$, *n.s.*) in female token versus all-male teams were not significantly different. However, token females engaged in more voice behavior ($M_{female} = 5.54$, $SD_{female} = 4.09$) compared to yellow-helmeted males in all male teams ($M_{male} = 3.51$, $SD_{male} = 3.19$; $F_{(1,155)} = 11.28$, $p < .05$).

Insert Table 1 about here.

Hypotheses 1a and 1b examined whether token female voice enactment in female token teams was lower or higher compared to that of a comparison male. To test Hypotheses 1a and 1b, we examined the direct effects of gender composition on yellow helmet voice enactment. Results showed that gender composition positively predicted yellow helmet voice enactment ($\gamma = .93$, $p < .05$; Model 2 of Table 2), indicating that even after controlling for voice frequency, token female voice enactment in female token teams was higher than a comparison male's voice enactment in an all-male team.[4] Thus, Hypothesis 1b (and not 1a) was supported.

Hypothesis 2 predicted leaders' gender beliefs would moderate the effect of gender on voice enactment, such that female tokens are likely to have higher voice enactment than a comparison male when leaders' gender beliefs are high than when they are low. As shown in Model 4 of Table 2, results indicated that gender composition interacted with leader's gender

---

[4] We also conducted ANOVA comparisons between the voice enactment of the token female and that of her male teammates wearing blue, green, or red helmets. We omitted comparisons with black-helmeted males because not all teams had a fifth member. Results revealed no significant differences between the token female's voice enactment and that of her male teammates.

beliefs to predict yellow helmet voice enactment ($\gamma = .56$, $p < .05$). Shown in Figure 1, leader gender beliefs was positively related to token female voice enactment in female token teams ($\gamma = .50$, $p < .05$) but unrelated to the yellow-helmeted male's voice enactment in all-male teams ($\gamma = -.02$, *n.s.*). This suggests that the presence of positive leader gender beliefs was essential for high token female voice enactment, but, interestingly, there appeared to be no penalty on token female voice enactment at low levels of leader gender beliefs. Moreover, these results were robust even when we removed cases where the token female was in the leadership role, suggesting that the effects of leader gender beliefs hold for male and female leaders. Lastly, leader gender beliefs had no effect on the yellow-helmeted male's voice enactment, suggesting that gender beliefs were uniquely activated in the presence of a token female, offering support for the validity of our gender beliefs measure in this context. Thus, Hypothesis 2 was supported.

Insert Table 2 and Figure 1 about here.

Hypothesis 3 predicted the relationship between token female voice enactment and team performance is contingent upon task complexity such that this relationship is positive when task complexity is high and negative when task complexity is low. As shown in Model 4 of Table 3, the three-way interaction term of team gender composition x yellow helmet voice enactment x task complexity was significant ($\gamma = 1.21$, $p < .05$). For ease of interpretation, in Figure 2, we split the interaction plot between yellow helmet voice enactment and task complexity into two graphs corresponding to female token teams versus all-male teams. Simple slopes analyses indicated that in female token teams, the relationship between token female voice enactment and team performance was positive and significant when task complexity was high ($\gamma = .50$, $p < .05$) but negative when task complexity was low ($\gamma = -.45$, $p < .05$). Additionally, results showed that

in all-male teams, yellow helmet voice enactment was unrelated to team performance when task complexity was high (γ = -.20, *n.s*) or low (γ = -.03, *n.s.*). Thus, our results provided support for Hypothesis 3. Notably, these relationships did not hold when we replaced token female voice enactment with voice frequency, suggesting that when linking her contributions to team outcomes, it is more important to focus on enacted voice as opposed to simply speaking up.[5]

Insert Table 3 and Figure 2 about here.

Supplementary Analyses

Although the above results provided support for our hypothesized model, we conducted several supplementary analyses to further understand the micro-dynamics of voice in diverse teams in our unique context (Hollenbeck & Wright, 2017).

The Impact of Voice Quality. One competing explanation for why the token female's voice was enacted is because her suggestions were expected to be or were of higher quality, and therefore facilitated team performance in complex tasks. We did not measure voice content or indicators of voice quality, but we tested this possibility by using survey data from female token teams that captured members' pre-task expectations prior to the team's attempts on any task ("To what degree do you expect this individual to offer valuable contributions helpful to the team's success?" *1 = not at all* to *7 = a very large extent*) and post-task perceptions of the quality of contributions for each of the four tasks ("During the task today, to what degree did this individual offer valuable contributions helpful to the team's success?" *1 = not at all* to *7 = a very large extent*). Although pre-task evaluations of the token female's expected contributions

---

[5] Following Aguinis, Gottfredson, and Joo (2013), we identified outliers in token female voice and voice enactment as +2.24 SD units or more above the mean within each task. 2 outliers pertained to voice enactment, and 9 outliers pertained to voice frequency. The results for Table 2 and Table 3 remained the same, despite excluding these outlier observations from the analyses.

(M_{female} = 4.68, SD = 1.39) was numerically lower than that of the yellow-helmeted male in all-male teams (M_{males} = 4.87, SD = 1.42), this difference was not significant ($F_{(1, 132)}$ = .49, *n.s.*). Similarly, the post-task evaluations of the token female (M_{female} = 5.37, SD = 1.53) was numerically lower than that of the yellow-helmeted male in all-male teams (M_{males} = 5.53, SD = 1.41), but this difference was not significant ($F_{(1, 531)}$ = 1.45, *n.s.*).

The Impact of Task Complexity. Another competing explanation is that the token female's voice was enacted more in complex tasks because more difficult and uncertain tasks increase majority members' openness to minority proposals, whereas in less complex tasks, majority members perceive minority input as creating unnecessary division. To explore this possibility, we examined whether task complexity interacted with the gender composition of the team to predict voice enactment. Regardless of gender composition, there was less yellow helmet voice enactment in high complexity tasks than in low complexity tasks (γ = -.82, $p < .05$), likely because it was harder to discern which suggestions should be enacted in highly complex tasks. ANOVA comparisons further revealed that token female voice enactment (M_{female} = 3.65) was higher than that of yellow helmeted-males (M_{male} = 1.65, $F_{(1,75)}$ = 9.69, $p < .05$) in low complexity tasks, but her voice enactment (M_{female} = 1.33) was not significantly different from yellow-helmeted males (M_{male} = 1.46, $F_{(1,75)}$ = .07*, n.s.)* in high complexity tasks. Ironically, this suggests that female voice was enacted more frequently in low complexity tasks (when enacting her voice appeared to harm team performance), but not more likely enacted in high complexity tasks (when enacting her voice helped performance).

The Impact of Leader's Gender Beliefs versus Team Beliefs. Although our hypothesis and results highlighted the impact of leader's gender beliefs on token female voice enactment in female token teams, it is also possible that fellow teammates' positive expectancies about

women's competence and capabilities (or even the token female's own positive beliefs) will facilitate her voice enactment. We amended Model 3 of Table 1 to include two interaction terms – one between non-yellow helmet beliefs and gender composition to reflect fellow male team members' beliefs, and one between gender composition and yellow helmet beliefs to reflect the token female member's beliefs. Neither interaction term significantly predicted yellow helmet voice enactment ($\gamma = .07$, *n.s.* and $\gamma = .29$, *n.s.*, respectively), indicating that above and beyond the leader's beliefs, neither the gender beliefs of male team members nor that of the token female explained additional variance in token female voice enactment in female token teams.

The Impact of Gender Composition on Deliberation. According to the minority influence literature, one explanation for the positive impact of token female voice enactment on complex tasks (and its negative impact on less complex tasks) is that minority voice produces more divergent and systematic processing in the team, which in turn slows down the information processing and decision-making of the team prior to taking action. To explore this possibility, we examined whether female token teams engaged in more voice before enacting a suggestion compared to all-male teams. Results showed that more suggestions occurred before the first enacted suggestion in female token teams (M = 3.34, SD = 3.70) compared to all-male teams (M = 1.63, SD = 2.99; $F_{(1, 84)} = 5.36$, $p < .05$). This corroborates theoretical arguments that exposure to minority viewpoints stimulates a wider search for information and the detection of multiple solutions (Nemeth & Rogers, 1996), and suggests that the lack of a minority viewpoint – especially in all-male teams because of their agentic nature – produced a proclivity toward action over more deliberative strategies. It seems, then, that in our context, the token female provided a "check and balance" on the tendency of all-male teams to engage in action-first strategies.

The Impact of Majority Member Voice. One might also conjecture about the role of male-majority (i.e., non-yellow helmet) voice on team performance in more or less complex tasks. The results of this analysis detected a significant three-way interaction between non-yellow helmet voice enactment, gender composition, and task complexity ($\gamma = -4.19$, $p < .05$; see Figure 3). Simple slopes analyses revealed that in female token teams, male majority voice enactment was *negatively* related to team performance when the task was complex ($\gamma = -.71$, $p < .05$) but positively related to team performance when the task was low in complexity ($\gamma = 2.05$, $p < .05$). This finding is the mirror opposite of the pattern we see in Figure 2 concerning token female voice enactment and consistent with research suggesting that majorities induced better performance in tasks where convergent thinking was useful and the perspective of the majority was appropriate (Nemeth, Mosier, & Chiles, 1992). Taken together, these results suggest that in female token teams, there may be something uniquely beneficial about enacting the token female's voice and something uniquely harmful about enacting the male majority's voice in high complexity tasks while the opposite holds true for low complexity tasks.

Insert Figure 3 about here.

The Impact of Physical Demand. Finally, given the highly contentious and polarized discussion about integrating women into all occupational categories in the military services, it was important to understand the role of physical demands in the performance of female token teams versus all-male teams. There is no shortage of evidence that men have more upper body strength and physical endurance compared to women, around which much of the debate regarding gender integration has centered (Michaels, 2013). We found a significant interaction effect between gender composition and task physicality in predicting team performance ($\gamma = -$

1.92, $p < .05$). This interaction indicated that female token teams and all-male teams performed similarly on low physicality tasks but that female token teams performed worse than all-male teams on high physicality tasks. Not surprisingly, when holding task complexity constant, there was a performance cost due to physical limitations associated with female token teams. We also examined whether female token teams performed differently than all-male teams in high-low combinations of task complexity and physicality. The three-way interaction between gender composition, task complexity, and task physicality was not significant, suggesting that the benefits of including a token female on complex tasks (and the costs of including her on physically demanding tasks) accrued in an additive rather than interactive manner.

DISCUSSION

Our findings reveal that token females had their voice enacted more than a comparison male in an all-male team, especially under a leader who espoused more positive gender beliefs. Moreover, token female voice enactment enhanced team performance in high complexity tasks but hurt team performance in low complexity, while male majority voice enactment exhibited the opposite effects. Our work offers several important theoretical and practical implications.

Theoretical Implications

Our first contribution is to go beyond voice frequency and highlight the importance of voice enactment as a construct linking the voice of individual members to team outcomes. Indeed, if one replaces our measure of "female voice enactment" with a measure of voice frequency, there is no relationship between gender composition and team performance even for complex tasks. The distinction between voice frequency and voice enactment is non-trivial. Although voice adds to the collective pool of ideas and has the potential to trigger new ideas, it does not appear to have direct implications for the team's performance outcomes. In comparison,

enacting the token female's voice represents a deeper integration of divergent courses of action into the team's overall approach to the task, which in turn facilitates team performance on complex tasks (van Knippenberg et al., 2004). Although the notion of voice enactment has been alluded to as one of many ways recipients can be receptive of voice (Detert & Burris, 2007; Fast et al., 2014), we extend this work by specifically studying its implications for team performance.

Our second contribution is to document *when* token women get their voice enacted. Although the literature on tokenism (Kanter, 1977) and gender stereotypes (Heilman, 2012) suggest that women will be disadvantaged when it comes to voice enactment, we show that female voice enactment can be higher than that of a comparison male's when leader's gender beliefs are high. To the extent leader's gender beliefs prevent attitudes consistent with tokenism and gender stereotypes from inhibiting token female voice enactment, and at the same time amplifying minority influence processes to increase token female voice enactment, this finding resolves contrasting predictions about token female voice enactment. Relatedly, our supplementary analyses revealed it was the leader's beliefs – and not that of the token female or the rest of the team – that were uniquely responsible for driving these effects. These findings affirm the influence of leaders and counter-stereotype beliefs in the team diversity literature (Fiske, Cuddy, Glick, & Xu, 2002; Mannix & Neale, 2005) and also highlight more broadly the importance of factors outside of women to facilitate female voice enactment and help move the needle toward gender parity.

Our third contribution is to clarify when enacting the token woman's voice benefits team performance. Our findings are consistent with past research suggesting that diverse teams can outperform homogeneous teams on complex tasks (van Knippenberg et al., 2004) and that greater female representation is beneficial for solving problems, including ones involving

realistic military scenarios (LePine, Hollenbeck, Ilgen, Colquitt, & Ellis, 2002; Hirschfeld, Jordan, Field, Giles, & Armenakis, 2005). Still, we extend this work by showing that it was token female voice enactment specifically that enhanced performance in cognitively demanding tasks. Identifying voice enactment and *whose* voice is enacted as a novel mechanism offers a new perspective on the processes underlying the effects of diversity on performance. In doing so, we address calls to investigate the micro-dynamics of diverse teams and elucidate *how* the behavior of certain members and the reactions they elicit can exert disproportionate effects on team outcomes for both good and ill (Emich & Wright, 2016; van Knippenberg & Mell, 2016).

Finally, we contribute to minority influence theory. Although some work has examined the efficacy of double minorities in influencing the way dominant members think about minority-relevant social issues (Kelley, 1971), it is unclear whether these effects generalize beyond experimental settings to affect the processes and performance of teams engaged in physically and cognitively demanding tasks (Maass & Clark, 1984). Relatedly, whereas the minority influence literature has focused on presentation tactics the minority individual should employ to increase his/her influence (Moscovici, 1980), our finding that leader gender beliefs and task complexity act as facilitators of both token female voice enactment and performance provides a broader understanding of factors outside of the minority individual that shape their impact on real team tasks. Taken together, our work demonstrates how minority influence theory can be extended and applied to address gender-based voice inequities in organizations.

Practical Implications

Our findings indicate that members and managers of female token teams need to pay attention to whose voice is enacted, because enacting the token female's (versus the male-majority's) voice exerts unique benefits (or costs) for team performance. If the task is high in

complexity, token females should speak up with unique ideas, and managers should seek to enhance the token female's voice enactment over that of the male majority because it may encourage the team to generate innovative performance strategies required to accomplish such tasks successfully. However, if the task is low in complexity and requires speedy completion, token females should focus on speaking up in ways that reflect the majority approach to efficiently reach a workable solution, and managers should be cautioned against enacting divergent voice as it may slow down the work.

Moreover, managers should recognize that merely having a token female present does not distinguish her team's performance from that of all-male teams. Rather, this woman has to speak up and, more importantly, her suggestions have to be enacted. A long history of gender diversity research would suggest that, due to tokenism, stereotype threat, and gender role expectations of women, female members may not speak up – and indeed, our field notes indicated that some of our female participants were passive, remained silent, or simply followed the directions of the other members. Yet, as our results show, when the token female did speak up, minority influence mechanisms appeared to work in favor of her voice enactment, and our findings suggest one way to further facilitate token female voice enactment is to select or train team leaders to hold more positive beliefs about women's capabilities in the relevant task domain.

As for practical implications in the military context specifically, those opposed to the integration of women focused much of their arguments on physical differences between the genders. Indeed, compared to all-male teams, female token teams performed worse on high physicality tasks, and our field notes indicated multiple instances where the token female was left behind in a physically demanding task because she could not independently execute a physical element of the task. Even so, the poorer performance exhibited by all-male teams on

high complexity tasks suggests that we need to look beyond a purely physical perspective to truly understand the pitfalls and opportunities of gender integration in the military. If recent history is any indication, many military mistakes that led to unnecessary loss of life or capture were more likely to result from poor tactical decision-making than from a lack of physical ability. Thus, the conversation regarding the liabilities of females in this context when it comes to physically demanding tasks should, at the very least, be complemented by a conversation of the severe liabilities of all-male teams in making tactical decisions on complex tasks.

Limitations and Directions for Future Research

We believe there are several noteworthy methodological strengths of this study. By manipulating gender composition and task complexity, randomly assigning individuals to teams, and rotating individuals through leadership roles, our within-subjects study design eliminated many potential alternative explanations that might obscure the effects of voice enactment on team performance. Moreover, our use of trained, reliable, and hypothesis-blind observers reduced biases that might be associated with retrospective self-reports, and observing all teams on all four tasks in a counter-balanced order mitigated confounds that might be attributable to fatigue or different stages of team development. Lastly, we examined representative individuals on representative tasks in a context in which gender-integration is a salient and contemporary issue. Taken together, these methodological strengths help to mitigate endogeneity concerns of self-selection and increase the internal and external validity of our findings.

Nevertheless, there are several key limitations in our work. Our sample contained very few women – twelve in all – who were recruited from a gender-integrated function to be token females in our teams. Although the small sample and functional background of these women was unavoidable given the context of our study, we are unable to rule out the possibility that there

was something particular about these women that may have advantaged their voice enactment in comparison to their male counterparts. Our supplementary analyses did not reveal systematic differences between the men and women in our sample with regards to expected or perceived quality of contributions, but it is possible that these women possessed higher competence (Rosette & Tost, 2010), cognitive ability, or specialized skills and training in their occupational function that contributed to their likelihood of voice enactment. Future work seeking to replicate our findings, thus, may increase the number of female participants, as well as control for individual differences in ability. As a related limitation, because our female and male participants were drawn from different departments, we acknowledge that our results may be in part driven by functional diversity rather than gender diversity. We think this is unlikely, because if our results were indeed driven by functional diversity, the moderating effects of leaders' gender beliefs would not have been expected. Still, we cannot completely rule out the possibility of functional diversity as an alternative explanation for our findings. Future work may improve upon ours by controlling for function when comparing all-male versus female token teams.

There is more work to be done on gender and voice enactment. Because this is the first study to examine voice enactment in teams, observing voice in real time during team task accomplishment was necessary to ensure an accurate interpretation of the phenomenon. However, this approach also prevented us from assessing gender or individual differences in *how* the voice was presented that may impact the likelihood of enactment. For example, men may present suggestions directly whereas women may present suggestions as questions (e.g., Maltz & Borker, 1982). Likewise, the minority influence literature suggests that consistency, confidence, and persistence in one's position can be important for influencing the majority (e.g., Moscovici, Lage, & Naffrechoux, 1969). Our field notes highlighted the relevance of these factors in one

instance where the token female proposed the correct solution to a complex task. Even though the rest of the team moved away from her suggestion, she persisted. Once the team eventually enacted her suggestion, the team was successful. We encourage future research to develop reliable instruments for assessing voice presentation. Future work may also track specific instances of voice and capture the quality of those suggestions to yield a more precise understanding of how enacting a particular suggestion facilitates team performance.

There are several context-specific factors that may have influenced our findings. Basing our four tasks on scenarios used by this institution to evaluate leadership and teamwork skills improved the ecological validity of our study, but we acknowledge that our task complexity manipulation may not have been as "controlled" as what one might encounter in a traditional laboratory experiment. Although we feel that trading ecological validity for comparability still yielded meaningful results, especially in view of our manipulation checks, we encourage future research to replicate our findings by utilizing standardized high and low complexity tasks, tasks with no physical component, or tasks where team outcomes are less time-sensitive. Moreover, given that teams in the military are characterized by a relatively high degree of authority differentiation (Hollenbeck, Beersma, & Schouten, 2012), the effect we documented for leaders' gender beliefs in the emergence of female voice enactment may be especially strong. Future research may investigate these relationships in contexts where authority differentiation is low.

Future research may also focus more on why leaders' gender beliefs matter. We theorized that these beliefs facilitate token female voice enactment because of more positive evaluations of the token female's voice, greater valuation of the uniqueness of her voice, or both, but we did not capture any of these mechanisms. Relatedly, we did not capture leaders' beliefs about the men's competence and capabilities, so it is unclear whether leaders' beliefs about women were as high

or higher than those about men and how those two sets of beliefs ultimately shape choices to enact the token female's versus the male majority's voice. Future research may investigate these possibilities and whether leader's gender beliefs produced specific behaviors, such as soliciting voice from token females, that increased the likelihood of her ideas being enacted.

Finally, future research may examine the role of voice enactment in teams marked by other forms of diversity. Given that a large number of occupations with a history of gender segregation are increasingly confronted with the realities of integration, we believe our findings on voice enactment extend existing research on gender diversity and team performance in meaningful ways. However, as diversity comes in many different forms, research on the possible effects of other types of diversity on voice enactment and team performance is also warranted. It may be that, depending on the type of diversity, majority versus minority voice enactment have different implications for team performance.

Conclusion

By integrating the employee voice, team diversity, and minority influence literatures, this study elucidates the unique role of token female voice enactment in traditionally male-dominated teams and identifies the conditions that facilitate its emergence and impact on team performance. Our work seeks to stimulate future study of not only voice in diverse teams but *whose* voice is enacted as an important mechanism affecting team outcomes. Our work also provides evidence-based and theoretically-backed guidance for managing voice enactment in female token teams. We hope it serves as a stepping stone to discovering the processes that enable all teams to excel.

REFERENCES

Aguinis, H., Gottfredson, R., & Joo, H. 2013. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods.,* 16(2): 270-301.

Argyris, C. 1977. Double loop learning in organizations. *Harvard Business Review,* 55(5): 115-129.

Asch, S. E. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs,* 70(9): 1-70.

Bales, R. F. 1950. *Interaction Process Analysis: A Method for the Study of Small Groups.* Cambridge, MA: Addison-Wesley.

Bauer, D. J., & Curran, P. J. 2005. Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research,* 40(3): 373-400.

Berg, J. M. 2014. The primal mark: How the beginning shapes the end in the development of creative ideas. *Organizational Behavior and Human Decision Processes,* 125: 1-17.

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

Bliese, P. D. 2000. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations:* 349-381. San Francisco: Jossey-Bass.

Boldry, J., Wood, W., & Kashy, D. A. 2001. Gender stereotypes and the evaluation of men and women in military training. *Journal of Social Issues,* 57: 689-705.

Bryk A. S., & Raudenbush, S. W. 1992. *Hierarchical linear models.* Newbury Park, CA: Sage.

Burris, E. R. 2012. The risks and rewards of speaking up: Managerial responses to employee voice. *Academy of Management Journal,* 55(4): 851-875.

Burris, E. R., Detert, J. R., & Romney, A. C. 2013. Speaking up vs. being heard: The disagreement around and outcomes of employee voice. *Organization Science,* 24: 22-38.

Cialdini, R. B., & Goldstein, N. J. 2004. Social influence: Compliance and conformity. *Annual Review of Psychology,* 55: 591-621.

Clark, R. D., & Maass, A. 1988. The role of social categorization and perceived source credibility in minority influence. *European Journal of Social Psychology,* 18: 381-394.

Correll, S. J., Ridgeway, C. L., 2006. Expectation States Theory. In: Delamater J. (Eds.) *Handbook of Social Psychology. Handbooks of Sociology and Social Research.* Springer, Boston, MA.

Crano, W. D., & Alvaro, E. M. 1998. Indirect minority influence: The leniency contract revisited. *Group Processes & Intergroup Relations,* 1(2): 99-115.

Crano, W. D., & Hannula-Bral, K. A. 1994. Context/categorization model of social influence: Minority and majority influence in the information of a novel response norm. *Journal of Experimental Social Psychology,* 30: 247-276.

De Dreu, C. K. W., & De Vries, N. K. (Eds.). 2001. *Group consensus and minority influence: Implications for innovation.* Oxford, UK: Blackwell Publishers Inc.

De Dreu, C. K., & Weingart, L. R. 2003. Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology,* 88: 741-749.

De Dreu, C. K., & West, M. A. 2001. Minority dissent and team innovation: The importance of participation in decision making. *Journal of Applied Psychology,* 86: 1191-1201.

DeRue, D. S., Nahrgang, J. D., Wellman, N., & Humphrey, S. E. 2011. Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. *Personnel Psychology,* 64(1): 7-52.

Detert, J. R., & Burris, E. R. 2007. Leadership behavior and employee voice: Is the door really open? *Academy of Management Journal,* 50: 869-884.

Deutsch, M., & Gerard, H. B. 1955. A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology,* 51(3): 629-636.

Eagly, A. H. 1987. Reporting sex differences. *American Psychologist,* 42: 756-757.

Eagly, A. H., & Johannesen-Schmidt, M. C. 2001. The leadership styles of women and men. *Journal of Social Issues,* 57: 781-797.

Eagly, A. H., & Wood, W. 1991. Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin,* 17: 306-315

Edmondson, A. C. 2003. Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies,* 40(6): 1419-1452.

Edmondson, A. C. 2012. Teamwork on the Fly. *Harvard Business Review,* 90(4): 72-80.

Edmondson, A. C., Bohmer, R. M., & Pisano, G. P. 2001. Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative Science Quarterly,* 46(4): 685-716.

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

Emich, K. J., & Wright, T. A. 2016. The 'I's in team: The importance of individual members to team success. *Organizational Dynamics,* 45(1): 2-10.

Enders, C. K., & Tofighi, D. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods,* 12(2): 121-138.

Enloe, C. 1983. *Does khaki become you? The militarization of women's lives.* Boston: South End Press.

Fairhurst, G. T., & Snavely, B. K. 1983. Majority and token minority group relationships: Power acquisition and communication. *Academy of Management Review,* 8(2): 292-300.

Farh, C. I. C., & Chen, G. 2018. Leadership and member voice in action teams: Test of a dynamic phase model. *Journal of Applied Psychology,* 103(1): 97-110.

Fast, N. J., Burris, E. R., & Bartel, C. A. 2014. Managing to stay in the dark: Managerial self-efficacy, ego defensiveness, and the aversion to employee voice. *Academy of Management Journal,* 57(4): 1013-1034.

Feingold, A. 1994. Gender differences in personality: A meta-analysis. *Psychological Bulletin,* 116(3): 429-456.

Finseraas, H., Johnsen, A. A., Kotsadam, A., & Torsvik, G. 2016. Exposure to female colleagues breaks the glass ceiling-evidence from a combined vignette and field experiment. *European Economic Review,* 90: 363-374.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology,* 82(6): 878-902.

Fitts, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Exponential Psychology,* 47: 381-391.

Forgas, J. P., & Williams, K. D. (Eds.). 2001. *Social influence: Direct and indirect processes.* Philadelphia; Psychology Press.

Gorenflo, D. W., & Crano, W. D. 1989. Judgmental subjectivity/objectivity and locus of choice in social comparison. *Journal of Personality and Social Psychology,* 57(4): 605-614.

Halpern, D. F. 1986. A different answer to the question, "Do sex related differences in spatial abilities exist?" *American Psychologist,* 41: 1014-1015.

Harrison, D.A., & Klein, K.J. 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review,* 32: 1199-1228.

Heilman, M. E. 2001. Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *The Journal of Social Issues,* 57(4): 657-674.

Heilman, M. E. 2012. Gender stereotypes and workplace bias. *Research in Organizational Behavior,* 32: 113-135.

Hirschfeld, R. R., Jordan, M. H., Feild, H. S., Giles, W. F., & Armenakis, A. A. 2005. Teams' female representation and perceived potency as inputs to team outcomes in a predominantly male field setting. *Personnel Psychology,* 58(4): 893-924.

Hofmann, D. A., & Gavin, M. B. 1998. Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management,* 24: 623-641.

Holland, S. L. 2006. The dangers of playing dress-up: Popular representations of Jessica Lynch and the controversy regarding women in combat. *Quarterly Journal of Speech,* 92: 27-50.

Hollenbeck, J. R., & Wright, P. M. 2017. Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management,* 43: 5-18.

Hollenbeck, J. R., Beersma, B., & Schouten, M. E. 2012. Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Academy of Management Review,* 37: 82-106.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. 1995. Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology,* 80: 292-316.

Howell, T. M., Harrison, D. A., Burris, E. R., & Detert, J. R. 2015. Who gets credit for input? Demographic and structural status cues in voice recognition. *Journal of Applied Psychology,* 100: 1765-1784.

Janis, I. 1972. *Victims of groupthink: A psychological study of foreign policy decisions and fiascoes.* Boston, MA: Houghton Mifflin.

Joshi, A., & Roh, H. 2009. The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal,* 52(3): 599-627.

Kanter, R. A. 1977. *Men and women of the corporation.* New York: Harper.

Kelley, H. H. 1971. Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior (pp. 1-26).* New York: General Learning Press.

Kozlowski, S. W. J., & Ilgen, D. R. 2006. Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest,* 7(3): 77-124.

LePine, J. A., Hollenbeck, J. R., Ilgen, D. R., Colquitt, J. A., & Ellis, A. 2002. Gender composition, situational strength, and team decision-making accuracy: A criterion decomposition approach. *Organizational Behavior and Human Decision Processes,* 88(1): 445-475.

LePine, J. A., & Van Dyne, L. 1998. Predicting voice behavior in work groups. *Journal of Applied Psychology,* 83: 853-868.

Li, A. N., Liao, H., Tangirala, S., & Firth, B. M. 2017. The content of the message matters: The differential effects of promotive and prohibitive team voice on team productivity and safety performance gains. *Journal of Applied Psychology,* 102(8): 1259-1270.

Liang, J., Farh, C. I., & Farh, J. L. 2012. Psychological antecedents of promotive and prohibitive voice: A two-wave examination. *Academy of Management Journal,* 55: 71-92.

Liang, J., Shu, R., & Farh, C. I. 2018. Differential implications of team member promotive and prohibitive voice on innovation performance in research and development project teams: A dialectic perspective. *Journal of Organizational Behavior.*

Lin, N. 1999. Social networks and status attainment. *Annual Review of Sociology,* 25(1): 467-487.

Lord, R. G. 1977. Functional leadership behavior: Measurement and relation to social power and leadership perceptions. *Administrative Science Quarterly,* 22(1): 114-133.

Loyd, D. L., Wang, C. S., Phillips, K. W., & Lount, Robert B., Jr. 2013. Social category diversity promotes premeeting elaboration: The role of relationship focus. *Organization Science,* 24(3): 757-772.

Maass, A., & Clark, R. D. 1984. Hidden impact of minorities: Fifteen years of minority influence research. *Psychological Bulletin,* 95(3): 428-450.

Maass, A., West, S. G., & Cialdini, R. B. 1987. Minority influence and conversion. *Review of Personality and Social Psychology,* 8: 55-79.

Maier, N. R., & Solem, A. R. 1952. The contribution of a discussion leader to the quality of group thinking: The effective use of minority opinions. *Human Relations,* 5: 277-288.

Maltz, D., & Borker, R. 1983. A cultural approach to male–female miscommunication. In J. Gumperz (Ed.), *Language and Social Identity (Studies in Interactional Sociolinguistics, pp. 196-216).* Cambridge: Cambridge University Press.

Mannix, E., & Neale, M. A. 2005. What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest,* 6(2): 31-55.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. 2001. A temporally based framework and taxonomy of team processes. *Academy of Management Review,* 26(3): 356-376.

Martin, R., Martin, P. Y., Smith, J. R., & Hewstone, M. 2007. Majority versus minority influence and prediction of behavioral intentions and behavior. *Journal of Experimental Social Psychology,* 43: 763-771.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. 1995. An integrative model of organizational trust. *Academy of Management Review,* 20: 709-734.

Maynes, T. D., & Podsakoff, P. M. 2014. Speaking more broadly: An examination of the nature, antecedents, and consequences of an expanded set of employee voice behaviors. *Journal of Applied Psychology,* 99: 87-112.

McClean, E., Martin, S. R., Emich, K. J., & Woodruff, T. 2018. The social consequences of voice: An examination of voice type and gender on status and subsequent leader emergence. *Academy of Management Journal,* 61: 1869-1891.

Mesmer-Magnus, J., & DeChurch, L. A. 2009. Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology,* 94(2): 535-546.

Michaels, J. 2013, February 3. Debate over women in combat shifts to physical strength. *USA Today.* Retrieved from http://www.usatoday.com/story/news/nation/2013/02/02/marines-army-standards-women-combat/1884141/

Morgeson, F. P., DeRue, D. S., Karam, E. P. 2010. Leadership in teams: A functional approach to understanding leadership structures and processes. *Journal of Management,* 36: 5-39.

Morgeson, F. P., & Humphrey, S. E. 2006. The Work Design Questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of Applied Psychology,* 91: 1321-1339.

Morrison, E. W. 2011. Employee voice behavior: Integration and directions for future research. *Academy of Management Annals,* 5: 373-412.

Morrison, E. W., & Milliken, F. J. 2000. Organizational silence: A barrier to change and development in a pluralistic world. *Academy of Management Review,* 25(4): 706-725.

Morrison, E. W., Wheeler-Smith, S., & Kamdar, D. 2011. Speaking up in groups: A cross-level study of group voice climate and voice. *Journal of Applied Psychology,* 96(1): 183-191.

Morrison, E. W. 2014. Employee voice and silence, *Annual Review of Organizational Psychology and Organizational Behavior,* 1: 173-197.

Moscovici, S. 1980. Toward a theory of conversion behavior. *Advances in Experimental Social Psychology,* 13: 209-239.

Moscovici, S. 1985. Social influence and conformity. In G. Lindzey, & E. Aronson (Eds.), *Handbook of social psychology (Vol. 2, pp. 347-412).* New York: McGraw-Hill.

Moscovici, S., Lage, E., & Naffrechoux, M. 1969. Influence of a consistent minority on the responses of a majority in a color perception task. *Sociometry,* 32(4): 365-380.

Moscovici, S., Mucchi-Faina, A., & Maass, A. (Eds.). 1994. *Minority influence.* Chicago: Nelson-Hall.

Moscovici, S., & Mugny, G. 1983. Minority influence. In P.B. Paulus (ED.), *Basic Group Processes,* (pp. 41-65). New York: Springer Verlag.

Moskowitz, G. B. 1996. The mediational effects of attributions and information processing in minority social influence. *The British Journal of Social Psychology,* 35: 47-66.

Mugny, G. 1982. *The power of minorities.* London: Academic Press.

Mugny, G., Butera, F., Sanchez-Mazas, M., & Pe´rez, J. A. 1995. Judgements in conflict: The conflict elaboration theory of social influence. In B. Booth, R. Hirsig, A. Helminger, B. Meier, & R. Volkart (Eds.), *Perception evaluation interpretation (pp. 160-168).* Gottingen: Hogrefe & Huber.

Nebus, J. 2006. Building collegial information networks: A theory of advice network generation. *Academy of Management Review,* 31: 615-637.

Nemeth, C. J. 1986. Differential contributions of majority and minority influence. *Psychological Review,* 93(1): 23-32.

Nemeth, C. J., & Kwan, J. L. 1985. Originality of word associations as a function of majority vs. minority influence. *Social Psychology Quarterly,* 48(3): 277-282.

Nemeth, C. J., & Kwan, J. L. 1987. Minority influence, divergent thinking and detection of correct solutions. *Journal of Applied Social Psychology,* 17(9): 788-799.

Nemeth, C. J., & Wachtler, J. 1983. Creative problem solving as a result of majority vs minority influence. *European Journal of Social Psychology,* 13(1), 45-55.

Nemeth, C. J., Connell, J. B., Rogers, J. D., & Brown, K. S. 2001. Improving decision making by means of dissent. *Journal of Applied Social Psychology,* 31(1): 48-58.

Nemeth, C. J., & Nemeth-Brown, B. 2003. Better than individuals? The potential benefits of dissent and diversity for group creativity. In P. B. Paulus, & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration; group creativity: Innovation through collaboration (pp. 63-84).* Oxford University Press, New York, NY.

Nemeth, C., & Rogers, J. 1996. Dissent and the search for information. *British Journal of Social Psychology,* 35(1): 67-76.

Nemeth, C., Mosier, K., & Chiles, C. 1992. When convergent thought improves performance: Majority versus minority influence. *Personality and Social Psychology Bulletin,* 18(2): 139-144.

Perlow, L., & Williams, S. 2003. Is silence killing your company? *Harvard Business Review,* 81: 52-58.

Phillips, K. W. 2003. The effects of categorically based expectations on minority influence: The importance of congruence. *Personality and Social Psychology Bulletin,* 29(1): 3-13.

Phillips, K. W., & Loyd, D. L. 2006. When surface and deep-level diversity collide: The effects on dissenting group members. *Organizational Behavior and Human Decision Processes,* 99(2): 143-160.

Phillips, K. W., Liljenquist, K. A., & Neale, M. A. 2009. Is the pain worth the gain? The advantages and liabilities of agreeing with socially distinct newcomers. *Personality and Social Psychology Bulletin,* 35(3): 336-350.

Phillips, K. W., Mannix, E. A., Neale, M. A., & Gruenfeld, D. H. 2004. Diverse groups and information sharing: The effects of congruent ties. *Journal of Experimental Social Psychology,* 40(4): 497-510.

Phillips, K. W., Northcraft, G. B., & Neale, M. A. 2006. Surface-level diversity and decision-making in groups: When does deep-level similarity help? *Group Processes & Intergroup Relations,* 9(4): 467-482.

Plaut, V. C. (2010). Diversity science: Why and how difference makes a difference. *Psychological Inquiry,* 21: 77-99.

Premeaux, S. F., & Bedeian, A. G. 2003. Breaking the silence: The moderating effects of self-monitoring in predicting speaking up in the workplace. *Journal of Management Studies,* 40(6): 1537-1562.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. 2004. *HLM 6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Ridgeway, C. L., & Correll, S. J. 2004. Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & Society,* 18(4): 510-531.

Rosette, A. S., & Tost, L. P. 2010. Agentic women and communal leadership: How role prescriptions confer advantage to top women leaders. *Journal of Applied Psychology,* 95(2): 221-235.

Rudman, L. A., & Glick, P. 2001. Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues,* 57(4): 743-762.

Rustad, Michael. 1982. *Women in Khaki: The American Enlisted Woman.* New York: Praeger.

Sandberg, S. 2013. *Lean in: Women, work, and the will to lead.* Random House.

Sandberg, S. & Grant, A. 2015, January 11. Speaking while female. *The New York Times.* Retrieved from https://www.nytimes.com/2015/01/11/opinion/sunday/speaking-while-female.html

Schwenk, C. R. 1990. Conflict in organizational decision making: An exploratory study of its effects in for-profit and not-for-profit organizations. *Management Science,* 36: 436-448.

Shrout, P. E., & Fleiss, J. L. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin,* 86(2): 420-428.

Statistic Brain. 2017. Women in the military statistics. Retrieved from https://www.statisticbrain.com/women-in-the-military-statistics/

Tessier, M. 2016, October 27. Speaking while female, and at disadvantage. *The New York Times.* Retrieved from https://www.nytimes.com/2016/10/27/upshot/speaking-while-female-and-at-a-disadvantage.html

Thomas-Hunt, M. C., Ogden, T. Y., & Neale, M. A., 2003. Who's really sharing? Effects of social and expert status on knowledge exchange within groups. *Management Science.* 49(4) 464-477.

U. S. Department of Labor. 1977. *Dictionary of occupational titles (4th ed.).* Washington, DC: U.S. Government Printing Office.

Volpato, C., Maass, A., Mucchi-Faina, A., & Vitti, E. 1990. Minority influence and social categorization. *European Journal of Social Psychology,* 20(2), 119-132.

van Dijk, H., Meyer, B., van Engen, M., & Loyd, D. 2016. Microdynamics in diverse teams: A review and integration of the diversity and stereotyping literatures. *Academy of Management Annals,* 11: 517-557.

van Knippenberg, D., De Dreu, C. K., & Homan, A. C. 2004. Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology,* 89: 1008-1022.

van Knippenberg, D., & Mell, J. N. 2016. Past, present, and potential future of team diversity research: From compositional diversity to emergent diversity. *Organizational Behavior and Human Decision Processes,* 136: 135-145.

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

Westphal, J. D., & Milton, L. P. (2000). How experience and network ties affect the influence of demographic minorities on corporate boards. *Administrative Science Quarterly,* 45(2): 366-398.

Woodworth, R. S. 1899. Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements,* 3(3): 1-114.

Yoder, J. D., Adams, J., & Prince, H. T. 1983. The price of a token. *Journal of Political and Military Sociology,* 11(2): 325-337.

Zaccaro, S. J., Rittman, A. L., & Marks, M. A. 2001. Team leadership. *Leadership Quarterly,* 12(4): 451-483.

TABLE 1

Bivariate Correlations between Study Variables

		M	SD	1	2	3	4	5	6	7	8	9
1	Task order	2.50	1.12									
2	Leader gender	0.08	0.27	-0.07								
3	Task physicality	0.50	0.50	-0.07	0.05							
4	Task complexity	0.50	0.50	-0.06	0.05	0.01						
5	Blue helmet voice frequency	5.08	3.90	0.02	-0.02	-0.01	0.23*					
6	Green helmet voice frequency	5.15	3.88	-0.01	-0.05	0.09	0.04	0.17*				
7	Red helmet voice frequency	4.47	3.53	0.06	-0.10	0.03	0.12	0.20*	-0.03			
8	Yellow helmet voice frequency	4.13	3.60	-0.08	0.20*	0.05	0.25*	0.30*	0.20*	0.06		
9	Black helmet voice frequency	4.36	2.49	0.07	-0.09	0.02	0.22*	0.49*	0.28*	0.40*	0.24*	
10	Blue helmet voice enactment	2.30	2.83	0.02	0.01	0.17*	-0.16†	0.71*	0.13†	0.18*	0.23*	0.35*
11	Green helmet voice enactment	2.75	3.14	-0.02	-0.09	0.27*	-0.30*	0.07	0.74*	-0.01	0.07	0.23*
12	Red helmet voice enactment	2.29	2.83	0.10	-0.07	0.09	-0.24*	0.09	-0.08	0.74*	-0.04	0.23*
13	Yellow helmet voice enactment	1.83	2.44	0.03	0.07	0.16*	-0.17*	0.14†	0.09	0.10	0.67*	0.16*
14	Black helmet voice enactment	2.13	1.92	0.06	-0.01	0.22*	-0.28*	0.25*	0.21*	0.34*	0.16*	0.65*
15	Leader's gender beliefs	4.38	1.40	0.06	0.19*	0.09	-0.05	0.01	0.21*	-0.05	0.07	0.06
16	Team performance	5.26	4.31	0.09	-0.03	-0.14†	-0.56*	-0.42*	-0.41*	-0.22*	-0.48*	-0.40*
17	Race diversity	0.61	0.24	-0.07	0.01	.	.	0.14	0.02	0.12	-0.06	0.34*
18	Rank diversity	0.63	0.19	0.03	0.16	.	.	-0.07	0.10	0.00	0.06	-0.18
19	Age diversity	1.95	1.00	0.15	-0.16	.	.	0.07	0.09	-0.08	-0.19	0.00
20	Gender composition	0.31	0.47	0.24	1.00*	.	.	-0.02	-0.05	-0.20	0.36*	-0.14
21	Team size	4.44	0.50	-0.14	0.07	.	.	-0.21	-0.19	-0.26	0.03	-0.47*

Note. N = 156 episodes in 39 teams; variables 1-16 are at the episodic level (Level 1); variables 17-21 are at the team level (Level 2).

The correlations between variables 1-16 with 17-21 are the correlations between averaged Level 1 variables and Level 2 variables.

Leader gender was coded as 1 = leader is female, 0 = leader is male.

[1] Coefficient alpha for leader gender beliefs.

*. Correlation is significant at the 0.05 level (two-tailed).

†. Correlation is significant at the 0.10 level (two-tailed).

TABLE 1 *(continued)*

Bivariate Correlations between Study Variables

		10	11	12	13	14	15	16	17	18	19	20
11	Green helmet voice enactment	0.28*										
12	Red helmet voice enactment	0.27*	0.11									
13	Yellow helmet voice enactment	0.35*	0.23*	0.23*								
14	Black helmet voice enactment	0.53*	0.52*	0.53*	0.42*							
15	Leader's gender beliefs	0.08	0.22*	0.03	0.10	0.13	(0.90)[1]					
16	Team performance	-0.10	-0.10	0.15†	-0.14†	0.05	-0.03					
17	Race diversity	0.02	0.11	0.11	-0.04	0.38*	0.49*	-0.09				
18	Rank diversity	0.01	-0.04	0.00	0.09	-0.19	0.17	-0.05	0.12			
19	Age diversity	0.15	0.20	-0.17	-0.25	0.005	0.04	-0.02	-0.17	0.08		
20	Gender composition	-0.06	-0.11	-0.07	0.30†	-0.09	0.28†	-0.07	0.02	0.17	-0.18	
21	Team size	-0.20	-0.26	-0.25	-0.06	-0.52*	-0.04	0.19	-0.32*	-0.02	-0.01	0.09

Note. $N = 156$ episodes in 39 teams; variables 1-16 are at the episodic level (Level 1); variables 17-21 are at the team level (Level 2).

The correlations between variables 1-16 with 17-21 are the correlations between averaged Level 1 variables and Level 2 variables.

Leader gender was coded as 1 = leader is female, 0 = leader is male.

[1] Coefficient alpha for leader gender beliefs.

*. Correlation is significant at the 0.05 level (two-tailed).

†. Correlation is significant at the 0.10 level (two-tailed).

TABLE 2

Regression Results Predicting Token Female (Yellow Helmet) Voice Enactment

	Yellow helmet voice enactment			
	M1	M2	M3	M4
Intercept	1.62**	1.45**	1.78	1.88
Episode level (Level 1)				
Task order	0.07	0.17	0.18	0.16
Leader gender	0.35	-0.68	-0.71	-1.02
Task physicality[1]	0.80**	0.65**	0.61**	0.71**
Task complexity[1]	-0.82*	-1.82**	-1.83**	-1.85**
Yellow helmet voice frequency[1]		0.57**	0.57**	0.56**
Leader's gender beliefs[1]			0.06	0.07
Team Level (Level 2)				
Team size[2]	-0.29	-0.33	-0.33	-0.32
Race diversity[2]	-0.75	-0.72	-0.57	-0.57
Rank diversity[2]	0.84	0.53	0.62	0.65
Age diversity[2]	-0.37*	-0.32*	-0.33*	-0.35*
Team gender composition[2]		0.93*	0.93*	0.99*
Leader's gender beliefs			-0.08	-0.08
Cross-Level Interactions				
Leader's gender beliefs[1] x Team gender composition[2]				0.56**
Pseudo R-squared	0.05	0.49	0.49	0.51
Model Deviance[3]	699.97	629.54	625.32	623.33

Note. $N = 156$ for Level 1 variables, $N = 39$ for Level 2 variables.

Team gender composition was coded 1 = female token teams, 0 = all-male teams.

[1] Group mean centered. [2] Grand mean centered.

[3] Deviance is a measure of model fit: it equals -2 x the log-likelihood of the maximum-likelihood estimate. The smaller the model deviance, the better the fit.
Pseudo R-squared values estimate the amount of total variance in the dependent variable captured by predictors in the model.

* $p < .05$ (one-tailed).   ** $p < .01$ (one-tailed).

TABLE 3

Regression Results Predicting Team Performance

| | | Team performance | | |
	M1	M2	M3	M4
Intercept	4.60**	8.40**	7.80**	7.96**
Episode level (Level 1)				
Task order	0.27	0.12	0.29*	0.28
Leader gender	0.06	-0.10	0.45	1.41
Task physicality[1]	-0.86*	-1.98**	-1.80**	-2.00**
Task complexity[1]			-2.23**	-2.18**
Blue helmet voice frequency[1]	-0.27*	-0.37**	-0.24**	-0.24*
Green helmet voice frequency[1]	-0.25**	-0.35**	-0.36**	-0.39**
Red helmet voice frequency[1]	-0.14	-0.41**	-0.24*	-0.23
Yellow helmet voice frequency[1]	-0.43**	-0.47**	-0.41**	-0.29**
Black helmet voice frequency[1]	-0.07	-0.40	-0.17	-0.19
Blue helmet voice enactment[1]		0.52**	0.36**	0.28
Green helmet voice enactment[1]		0.09	0.03	0.14
Red helmet voice enactment[1]		0.29*	0.06	0.05
Yellow helmet voice enactment[1]		-0.07	-0.05	-0.09
Black helmet voice enactment[1]		0.65*	0.37	0.63*
Blue helmet voice enactment[1] x Task complexity[1]			-0.41	-0.59*
Green helmet voice enactment[1] x Task complexity[1]			-0.24	-0.32
Red helmet voice enactment[1] x Task complexity[1]			0.29	-0.20
Yellow helmet voice enactment[1] x Task complexity[1]			-0.28	0.03
Black helmet voice enactment[1] x Task complexity[1]			1.71**	1.17*

Note. $N = 156$ for level 1 variables, $N = 39$ for level 2 variables.
Team gender composition was coded 1 = female token teams, 0 = all-male teams.
[1] Group mean centered. [2] Grand mean centered. [3] Deviance is a measure of model fit: it equals -2 × the log-likelihood
 of the maximum-likelihood estimate. The smaller the model deviance, the better the fit.
 Pseudo R-squared values estimate the amount of total variance in the dependent variable captured by predictors in the model.
* $p < .05$ (one-tailed).   ** $p < .01$ (one-tailed).

TABLE 3 (*continued*)

Regression Results Predicting Team Performance

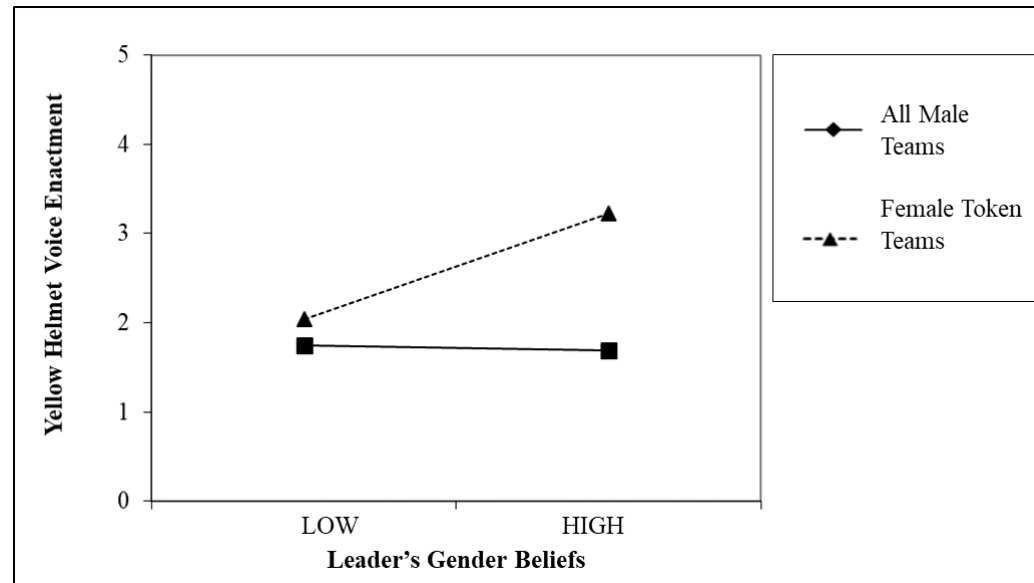|	Team performance			
	M1	M2	M3	M4
Team Level (Level 2)				
Team size[2]	0.82	0.10	0.03	-0.07
Race diversity[2]	-0.42	-1.08	-1.32	0.31
Rank diversity[2]	-0.48	0.07	-0.05	-0.72
Age diversity[2]	-0.04	-0.01	0.08	0.24
Blue helmet voice enactment		-0.46**	-0.49**	-0.47*
Green helmet voice enactment		-0.46*	-0.46*	-0.49*
Red helmet voice enactment		0.06	-0.06	0.05
Yellow helmet voice enactment		-0.78**	-0.96**	-0.68*
Black helmet voice enactment		0.10	0.50	0.16
Gender composition[2]			-0.14	-0.65
Cross-Level Interactions				
Task complexity[1] x Gender composition[2]			1.64**	2.92**
Blue helmet voice enactment[1] x Gender composition[2]			0.02	0.16
Green helmet voice enactment[1] x Gender composition[2]			0.06	0.52*
Red helmet voice enactment[1] x Gender composition[2]			-0.02	-0.10
Yellow helmet voice enactment[1] x Gender composition[2]			-0.11	0.14
Black helmet voice enactment[1] x Gender composition[2]			-0.49	0.07
Blue helmet voice enactment[1] x Task complexity[1] x Gender composition[2]				-2.03**
Green helmet voice enactment[1] x Task complexity[1] x Gender composition[2]				-0.77
Red helmet voice enactment[1] x Task complexity[1] x Gender composition[2]				-1.77**
Yellow helmet voice enactment[1] x Task complexity[1] x Gender composition[2]				1.21*
Black helmet voice enactment[1] x Task complexity[1] x Gender composition[2]				0.57
Pseudo R-squared	0.29	0.62	0.67	0.70
Model Deviance[3]	841.78	765.78	740.17	716.64

FIGURE 1

Effect of Leader's Gender Beliefs on Token Female (Yellow Helmet) Voice Enactment

FIGURE 2

Moderating Effect of Task Complexity on the Relationship between Token Female (Yellow Helmet) Voice Enactment and Team Performance, in All-Male Teams compared to Female Token Teams
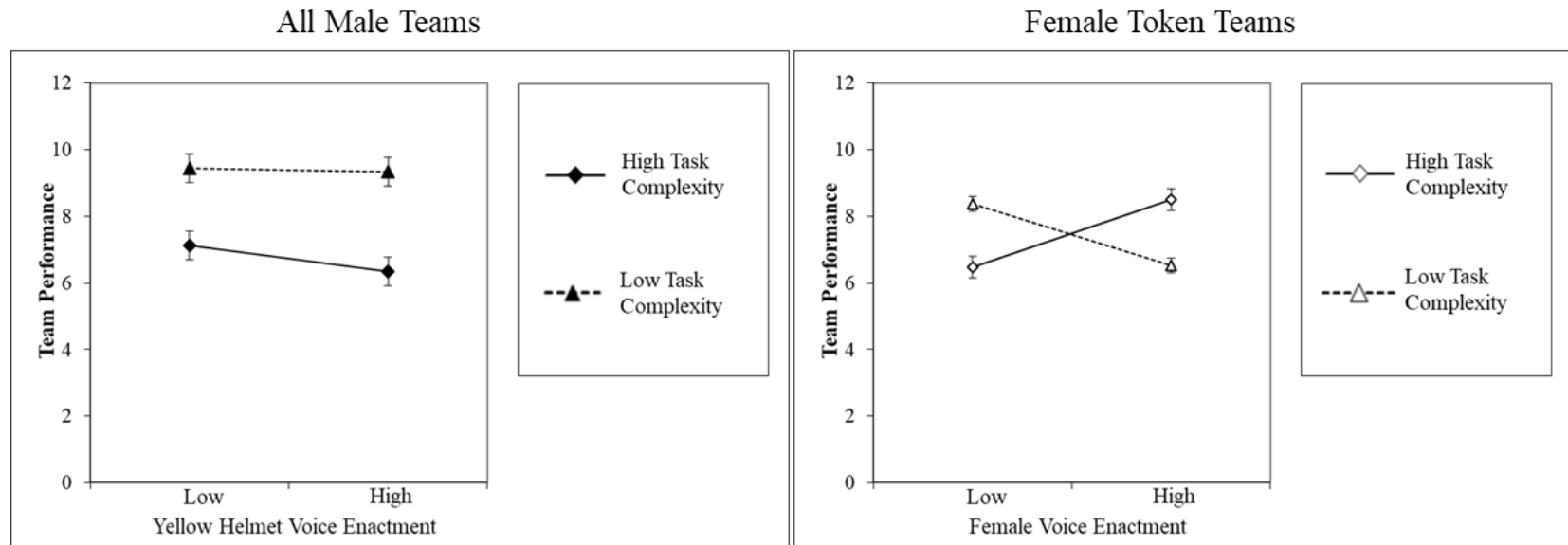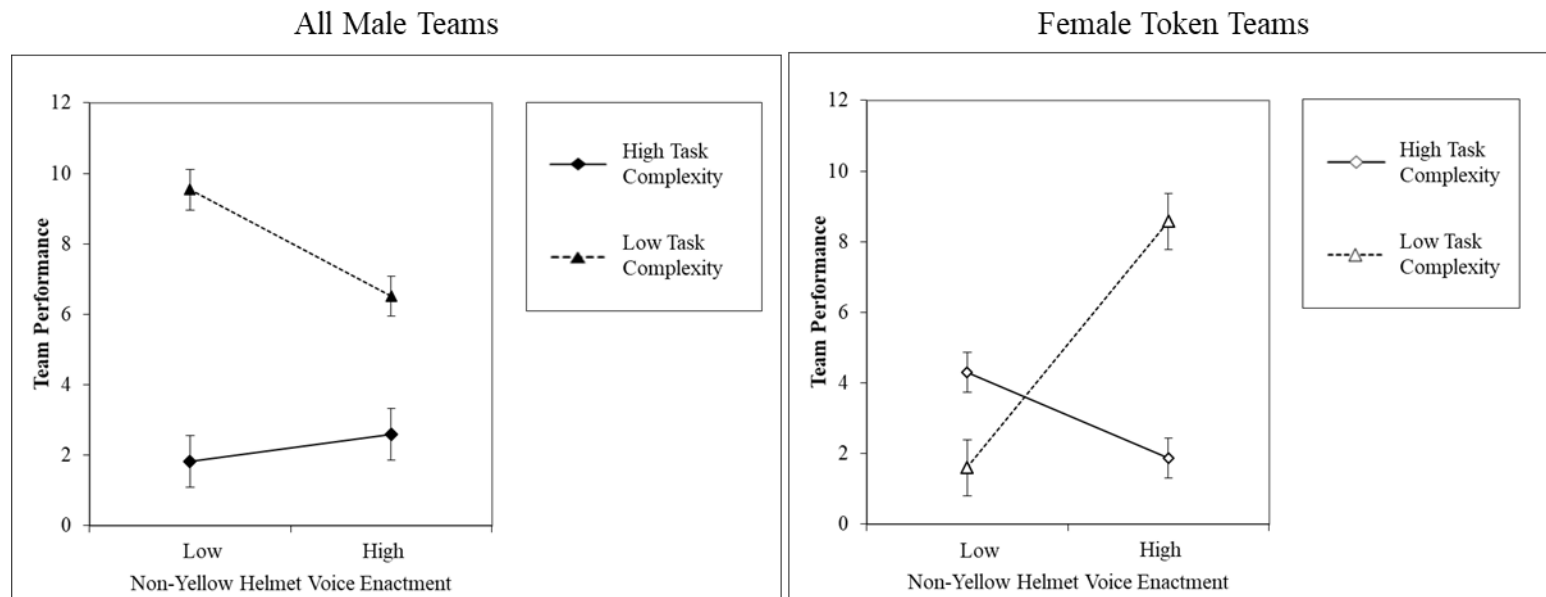
FIGURE 3

Supplementary Analysis: Moderating Effect of Task Complexity on the Relationship between Non-Yellow Helmet Voice Enactment and Team Performance, in All-Male Teams compared to Female Token Teams

APPENDIX

Task A. Casualty Rescue

Participants use the wooden plank to move a dummy through a hole in the wall.

Task B. Building Resupply

Participants climb onto a pedestal by using the wooden plank as a bridge.

Task C. Medicine Delivery

Participants devise a pulley system to send "medicine" across an off-limits zone.

Task D. Chemical Walls

Participants access the "safe" zone on top of the walls by building a platform.