

1
2
3
4
5
6
7
8
9

Token Female Voice Enactment in Traditionally Male-Dominated Teams: Facilitating Conditions and Consequences for Performance

10
11
12

Crystal I. C. Farh

University of Washington
farh@uw.edu

13
14

Jo K. Oh

University of Connecticut
kyoungjo.oh@uconn.edu

15
16
17
18

John R. Hollenbeck

Michigan State University
jrh@broad.msu.edu

19
20
21
22

Andrew Yu

University of Melbourne
andrew.yu@unimelb.edu.au

23
24
25
26

Stephanie M. Lee

Baylor University
stephanie_kunst@baylor.edu

27
28
29
30
31

Danielle D. King

Rice University
danielle.d.king@rice.edu

32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

In press at *Academy of Management Journal*

49

Acknowledgements

50
51
52
53
54
55
56
57
58
59
60

We thank associate editor Lisa Leslie and three anonymous reviewers for their extremely thorough and constructive guidance throughout the review process. We are also indebted to Larry Farh, Daniel Feiler, Laura Huang, Jennifer Marr, and the participants of the University of Washington M&O brown bag series and University of Pennsylvania Wharton OB conference for their very helpful comments and insights on earlier drafts of this paper. Finally, we thank George Akst, Jared Gastrock, and Juan Iturriaga for their invaluable assistance in the data collection process.

1
2
3 **TOKEN FEMALE VOICE ENACTMENT IN TRADITIONALLY MALE-DOMINATED**
4 **TEAMS: FACILITATING CONDITIONS AND CONSEQUENCES FOR**
5 **PERFORMANCE**
6
7
8

9 **ABSTRACT**
10

11 When is a token female's voice incorporated into the actions of a traditionally male-
12 dominated team and to what ends? Drawing from the tokenism, gender stereotypes, and minority
13 influence literatures, we advance a model that specifies the conditions that facilitate token female
14 voice enactment and when enacting her voice enhances team performance. Using a sample of
15 active duty military men and women, we employed live observation techniques to study voice
16 enactment in all-male teams versus female token teams (i.e., teams with a token female member)
17 throughout a series of complex and physically demanding tasks. Our findings revealed that a)
18 token female voice enactment was higher when team leaders possessed more favorable beliefs
19 about women's capabilities in the military, and b) token female voice enactment enhanced team
20 performance in more complex tasks but harmed team performance in less complex tasks.
21 Additionally, our supplementary analyses revealed that female token teams were more reflective
22 before engaging in action relative to all-male teams that tended to engage in agentic, "action-
23 first" strategies. Theoretical and practical implications for facilitating female voice enactment in
24 traditionally male-dominated contexts are discussed.
25
26
27

28 *Keywords:* Gender diversity; Leadership; Minority influence; Teams; Tokenism; Voice; Voice
29 enactment
30
31
32

33 Employee voice – defined as individuals' discretionary communication of work-related
34 ideas, suggestions, concerns, or opinions (Morrison, 2011) – is believed to be important for
35 promoting effective team functioning. By converting private knowledge held by individual
36 members into public knowledge for the team to use, voice ensures teams a) have up-to-date
37 relevant task information, b) generate multiple decision alternatives, and c) engage in adaptive
38 coordination in the face of change (Edmondson, 2003, 2012; Edmondson, Bohmer, & Pisano,
39 2001). Nevertheless, there are obstacles that threaten to nullify these benefits for team
40 performance. Voice may not always be expressed (Morrison & Milliken, 2000), recognized
41 (Burriss, Detert, & Romney, 2013), or positively evaluated (Burriss, 2012). Moreover, once
42 expressed, voice may not necessarily be acted upon (Fast, Burriss, & Bartel, 2014). These issues
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 may be particularly salient for minority members who differ from the majority on observable
4 socio-demographic features and encounter difficulties in getting their voice enacted by the team.
5
6

7
8 Take, for example, women who are newly integrated into traditionally male-dominated
9 settings, such as the military, top management, science, and engineering (Joshi & Roh, 2009).
10 Anecdotal evidence suggests that when it comes to voice and voice enactment in these contexts,
11 the token female – defined as the solo female in a group of males (Kanter, 1977) – may find her
12 voice ignored or “drowned out” (Sandberg & Grant, 2015; Tessier, 2016). If a diverse workforce
13 generates more performance-enhancing alternatives than homogeneous teams (De Dreu & West,
14 2001), there is clearly a need to understand when women’s voices – especially that of a token
15 female in an all-male team – are enacted. There is also a need to understand when enacting her
16 voice facilitates team performance, as this may reveal the costs of failing to do so.
17
18
19
20
21
22
23
24
25
26
27

28 In the voice literature, recent work has begun to explore gender differences in whose
29 voice is recognized and given credit (Howell, Harrison, Burriss, & Detert, 2015; McClean,
30 Martin, Emich, & Woodruff, 2018), but this research has primarily focused on women’s voice
31 frequency rather than voice enactment – i.e., its incorporation into real-time decisions related to
32 team strategies and processes – and its potential impact on team outcomes. In the diversity
33 literature, traditional perspectives suggest token females are evaluated negatively by majorities
34 due to a host of perceptual, attributional, and stereotype-driven processes (Kanter, 1977;
35 Heilman, 2012). However, other research suggests that the presence of a demographically
36 dissimilar other can increase a team’s openness to different perspectives (Phillips, Liljenquist, &
37 Neale, 2009; Phillips, Northcraft, & Neale, 2006). These competing rationales appear to present
38 mutually exclusive arguments for predicting token female voice enactment – and neither
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 perspective describes the factors that may tip forces in favor of her voice enactment nor the
4
5 conditions under which enacting her voice benefits team performance.
6
7

8 To address these limitations, our theoretical model draws on the tokenism, gender
9
10 stereotype, and minority influence literatures to answer two related research questions. First,
11
12 what conditions facilitate token female voice enactment in traditionally male-dominated team
13
14 contexts? Second, what are the implications of enacting token female voice for team
15
16 performance? As token females are often “new minorities” in otherwise all-male settings, team
17
18 members’ may not have prior experiences to guide interactions with the token female, and her
19
20 voice is likely to be evaluated on the basis of social-group expectations informed by stereotypes.
21
22 Additionally, because token females not only look different but are also likely to present
23
24 viewpoints that differ from those of an all-male majority (Eagly & Wood, 1991; Feingold, 1994;
25
26 Halpern, 1986), we draw on the minority influence literature to understand when a
27
28 demographically dissimilar individual can shape the thoughts and actions of the majority
29
30 (Moscovici, 1980; Nemeth, 1986; Phillips et al., 2006; Thomas-Hunt, Ogden, & Neale, 2003).
31
32 By integrating and extrapolating from these perspectives, we further identify leader’s gender
33
34 beliefs as a facilitator of token female voice enactment and task complexity as a boundary
35
36 condition of when enacting her voice will facilitate team performance in female token teams.
37
38
39
40
41

42 Our work offers several contributions. We extend voice research by going beyond
43
44 investigations of voice frequency (e.g., Morrison, 2014) to focus on voice enactment, a construct
45
46 alluded to as an indicator of managerial receptivity (Fast et al., 2014) but not yet formally studied
47
48 in the literature. By examining the conditions that facilitate token female voice enactment and
49
50 the performance implications of acting on her voice, we offer new insight on when the voice of
51
52 demographic minorities can impact and shape performance in majority-dominated team contexts
53
54
55
56
57
58
59
60

1
2
3 (Howell et al., 2015). We also extend the team diversity literature by identifying boundary
4 conditions, such as the leader's gender beliefs, that reconcile competing perspectives around
5 whether and when token females' voices are enacted. More broadly, our focus on voice
6 enactment investigates the micro-dynamics that enable specific individuals to affect the
7 performance outcomes in diverse teams (van Knippenberg & Mell, 2016; van Dijk, Meyer, van
8 Engen, & Loyd, 2016) and sheds light on the processes that underlie contingency factors of
9 diversity effects (e.g., Plaut, 2010; van Knippenberg & Schippers, 2007). Lastly, we extend the
10 minority influence literature, which has traditionally focused on how those expressing minority
11 opinions produce conversion among the majority (Moscovici, 1980) but rarely examined when
12 the opinions of newly integrated, socio-demographic tokens are enacted and shape performance.
13
14
15
16
17
18
19
20
21
22
23
24
25

26 Beyond these theoretical extensions, our work presents an important contribution because
27 we test our model in a branch of the U.S. military that, at the time of data collection, was
28 undergoing legislative changes to open certain all-male job categories to women for the first time
29 in history. Women were a new and unprecedented minority in this context, and the male majority
30 had limited previous experience working with them on interdependent tasks (Joshi & Roh,
31 2009). Thus, our work offers evidence-guided and timely implications for managing newly
32 integrated teams in a setting of national importance.
33
34
35
36
37
38
39
40
41

42 **THEORETICAL BACKGROUND**

43 **Voice and Voice Enactment in Teams**

44 In team settings, members exhibit voice when they speak up with improvement-oriented
45 suggestions, concerns, and opinions related to the team's task (Farh & Chen, 2018; Li, Liao,
46 Tangirala, & Firth, 2017; Morrison, Wheeler-Smith, & Kamdar, 2011). Unlike in supervisor-
47 subordinate dyads, voice in teams is often directed towards the team leader *as well as* fellow
48 team members because teams are characterized by interdependence, diffusion of expertise, and
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 distributed authority for action (Li et al., 2017; Morrison et al., 2011; Morrison, 2014).
4
5 Moreover, because team interactions revolve around the accomplishment of dynamic team tasks
6
7 (Kozlowski & Ilgen, 2006; Marks, Mathieu, & Zaccaro, 2001), the status quo upon which voice
8
9 attempts to improve reflects the team's current reasoning or approach to the team task (Premeaux
10
11 & Bedeian, 2003). Thus, in theory, voice helps the team iteratively improve its current approach
12
13 to accomplishing its tasks (Edmondson, 2003; Edmondson et al., 2001; Farh & Chen, 2018).
14
15
16

17 Voice is considered an essential source of health and functioning for teams (Mesmer-
18
19 Magnus & DeChurch, 2009; Nemeth, Connell, Rogers, & Brown, 2001), and when members fail
20
21 to exercise voice, there can be serious negative consequences for team effectiveness and
22
23 performance (Argyris, 1977; Janis, 1972; Perlow & Williams, 2003). Yet, scholars are
24
25 increasingly realizing that simply engaging in voice, does not guarantee that one's suggestions
26
27 will impact the team in consequential ways. For instance, ignoring or contradicting a member's
28
29 voice limits its impact. Likewise, simply valuing voice does not imply action. On the other hand,
30
31 *voice enactment* – which we conceptualize as the extent voice is incorporated into the team's
32
33 approach to the task – has the potential to directly shape team outcomes.
34
35
36
37

38 Because voice enactment requires voice recipients to take action on the speaker's
39
40 suggestion, it reflects a type of interpersonal influence (Correll & Ridgeway, 2006). Still, voice
41
42 enactment can be distinguished from other types of influence that reflect changes in opinion,
43
44 thought, or behavior that may or may not have occurred in response to voice, such as conformity
45
46 to social norms or obedience to authority (Cialdini & Goldstein, 2004). Voice enactment is also
47
48 distinct from informal leadership, which occurs when a member who is not the formal leader acts
49
50 to satisfy team needs and enhance team effectiveness (Morgeson, DeRue, & Karam, 2010).
51
52
53
54 Though a team member's voice is a type of change-oriented behavior (DeRue, Nahrgang,
55
56
57
58
59
60

1
2
3 Wellman, & Humphrey, 2011), informal leadership also encompasses non-voice forms of
4 influence, such as role modeling or building a positive socioemotional team climate (Lord,
5 1977). Thus, we prefer the label “voice enactment” because it more precisely captures the
6 change-oriented manner in which a member influences the team toward team goals.
7
8
9

10
11
12 Given our focus on voice enactment and its implications for team performance, we make
13 two assumptions going forward. First, although voice can encompass suggestions, concerns, and
14 other opinions (Liang, Farh, & Farh, 2012; Maynes & Podsakoff, 2014), we focus on suggestion-
15 based forms of voice, which convey proposed actions and alternatives (LePine & Van Dyne,
16 1998; Premeaux & Bedeian, 2003) that – if enacted by the team – can have direct and immediate
17 implications for the team’s task accomplishment (Hollenbeck, Ilgen, Sego, Hedlund, Major, &
18 Phillips, 1995). Opinions or concerns may play an important role in shaping team processes
19 (Liang, Shu, & Farh, 2018), but they do not always present actionable ideas. Second, we
20 acknowledge that voice enactment requires the speaker to first express voice. That is, voice not
21 expressed cannot be enacted. Thus, while potential gender differences in voice frequency
22 (Sandberg & Grant, 2015) may have downstream consequences for voice enactment, the focus of
23 our work is to investigate what happens to token females’ voice *after* it is expressed.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 **Token Female Voice Enactment in Male-Dominated Teams**

41
42 To understand how a homogeneous male majority may react to token female voice, we
43 integrate perspectives from the tokenism, gender stereotypes, and minority influence literatures.
44 Tokens refer to individuals who are identified by an ascribed characteristic that is shared by few
45 others in the group and carries a set of assumptions about culture, status, and behavior (Kanter,
46 1977). Because tokens are lone representatives of their ascribed category, their perceived
47 differences from the dominant group tend to be exaggerated by majority members, and their
48 actions are often scrutinized through the lens of stereotypes (Rustad, 1982; Yoder, Adams, &
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Prince, 1983). This is likely in the case of token females, whose lower status in society's sex
4 hierarchy reduces their power to defy a male majority's assertions of gender stereotypes
5
6 (Fairhurst & Snavely, 1983). Likewise, in contexts where token females are "newly" integrated
7
8 into male-dominated settings, interpersonal knowledge and experience has not yet been
9
10 established. Thus, gender stereotypes likely guide the male majority's initial perceptions of and
11
12 interactions with the token female (Finseraas, Johnsen, Kotsadam, & Torsvik, 2016).
13
14
15

16
17 The gender stereotype literature indicates how a number of cognitive and attributional
18 forces may work against token female voice enactment. First, descriptive and prescriptive gender
19 stereotypes imply that women are and should be communal, conforming, nurturing, and self-
20 effacing (Heilman, 2012). To the extent the male majority expects token females to conform to
21 these pre-existing generalizations about their category (Kanter, 1977), they are likely to perceive
22 her engagement in an agentic, change-oriented behavior such as voice as incongruent with their
23 expectations (Eagly, 1987; McClean et al., 2018). Thus, women may incur penalties for voicing
24 their suggestions or have those suggestions negatively evaluated by the male majority (Heilman,
25 2001; Rudman & Glick, 2001). Second, the hierarchical dimension of gender stereotypes
26 produces expectations that women are less competent than men (Ridgeway & Correll, 2004),
27 which in turn may lead male counterparts to ignore or dismiss her suggestions, deeming them as
28 unlikely to benefit the team. Finally, because being a token exaggerates perceived differences
29 between the dominant male majority and the token female, efforts may be made to socially
30 isolate her as an outsider, further reducing the likelihood of trusting and enacting her voice
31 (Kanter, 1977; Yoder et al., 1983). For these reasons, it is plausible that token female voice
32 enactment will be low in traditionally male-dominated teams.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 In contrast, the literature on minority influence suggests that perceived dissimilarity from
4 a minority individual can facilitate majority openness to the minority's perspectives.
5
6 Specifically, research on minority influence (e.g., De Dreu & De Vries, 2001; Moscovici,
7
8 Mucchi-Faina, & Maass, 1994) argues that when an individual presents a view that is perceived
9
10 as different from the rest of the group, members of the majority are *more* likely to process the
11
12 content of his or her view in a systematic way. That is, although the majority may be motivated
13
14 to reject the minority position (Asch, 1956; Deutsch & Gerard, 1955; Nemeth, 1986), the act of
15
16 resisting it leads majority individuals to carefully evaluate the arguments raised. In turn, this
17
18 systematic processing illuminates the true merits of the message, produces a deeper
19
20 understanding of the issue, and ultimately increases the influence of the minority (Forgas &
21
22 Williams, 2001; Maass, West, & Cialdini, 1987; Mugny, 1982; Mugny, Butera, Sanchez-Mazas,
23
24 & Perez, 1995; Nemeth & Wachtler, 1983). Moreover, there is research suggesting that
25
26 attitudinal changes due to minority influence tend to be accompanied by changes in behavioral
27
28 intentions (Martin, Martin, Smith, & Hewstone, 2007), suggesting an increased likelihood of
29
30 enacting a minority member's voice.
31
32
33
34
35
36
37

38 Although most of the above work was conducted in relation to minority opinion holders,
39
40 some research suggests that the basic mechanism of increased systematic processing generalizes
41
42 to individuals known as "double minorities" (i.e., those who also differ from the majority in
43
44 terms of their social category; Moscovici, 1985). For example, Phillips (2003) argued that a
45
46 minority opinion of a social category outgroup member may engender greater information
47
48 processing in the team compared to if that opinion came from an ingroup member (Crano &
49
50 Avaro, 1998; Crano & Hannula-Bral, 1994; Gorenflo & Crano, 1989). This is because majority
51
52 members expect other ingroup members to hold similar opinions to their own. As such, a deviant
53
54
55
56
57
58
59
60

1
2
3 opinion from an ingroup member produces a certain level of dissonance or relational discomfort
4
5 and is met with negative evaluation. In contrast, because majority members are less concerned
6
7 about their social relationship with an outgroup member and do not expect categorically
8
9 dissimilar others to share the majority's perspectives and opinions, exposure to a minority
10
11 individual's dissenting opinion generates less discomfort that may otherwise interfere with
12
13 processing the minority perspective (Phillips, 2003). In fact, the presence of categorically
14
15 dissimilar others may heighten the group's awareness that demographic minority members have
16
17 different information or ideas that may potentially benefit the team's task accomplishment,
18
19 which in turn may legitimize unique perspectives voiced by a minority member (Phillips, 2003).
20
21
22
23

24 Indirectly supporting these claims, research has shown that when teams contained a
25
26 demographically dissimilar individual, the team was more likely to make use of informational
27
28 diversity (Phillips, 2003; Phillips, Mannix, Neale, & Gruenfeld, 2004). Likewise, Thomas-Hunt
29
30 et al. (2003) found that socially connected ingroup members gave greater attention to the unique
31
32 information contributed by socially isolated outgroup members than to information contributed
33
34 by other ingroup members. Research has also shown that groups with greater surface-level
35
36 diversity (e.g., age, gender, race) exhibited more task engagement, shared more information, and
37
38 expressed more dissenting opinions compared to homogeneous groups (Loyd, Wang, Phillips, &
39
40 Lount, 2013; Phillips & Loyd, 2006; Phillips et al., 2006). Although none of this work looked
41
42 specifically at how these effects translate into token female voice enactment in an otherwise all-
43
44 male team, they nonetheless provide a counter-narrative in which a token female is likely to have
45
46 her voice enacted, precisely because of her token status and minority perspective.
47
48
49
50

51 Taken together, contrasting forces appear to shape the male majority's likelihood of
52
53 enacting a token female's voice. On the one hand, the majority's perceptions and attributions of a
54
55
56
57
58
59
60

1
2
3 stereotyped token female may produce negative evaluations that render her voice enactment less
4 likely. On the other hand, expectations of unique information guided by minority influence
5 mechanisms may cause the male majority to engage in greater information processing and
6 consideration of a token female's suggestions, making her voice enactment more likely. To
7 reflect these divergent predictions in the literature, we offer the following competing hypotheses:
8
9

10
11
12
13
14
15 *Hypothesis 1a: Token female voice enactment will be lower than a comparison male.*

16
17 *Hypothesis 1b: Token female voice enactment will be higher than a comparison male.*

18 19 **Leader's Gender Beliefs as a Facilitator of Token Female Voice Enactment**

20
21 We now turn to consider factors that likely turn the tide in favor of token female voice
22 enactment. It seems that if factors exist to counteract negative perceptions, attributions, and
23 stereotypes about female members and at the same time legitimize the value of systematically
24 processing her view points, token female voice enactment is more likely to occur (Ely &
25 Thomas, 2001). One such factor may be favorable gender beliefs. In particular, beliefs that
26 women are competent and capable in the relevant work domain or occupation are most pertinent
27 for facilitating token female voice enactment because they counteract the status-based gender
28 stereotype that women are less competent than men (Ridgeway & Correll, 2004). Indeed,
29 research suggests that when perceivers possess counter-stereotype expectations about an actor's
30 social group, normative fit is reduced, and they are less likely to categorize the actor based on his
31 or her social categories (van Knippenberg, De Dreu, & Homan, 2004). At the same time, the
32 minority influence literature suggests that when the majority holds positive attributions of the
33 minority member, majority members are more likely to systematically process the message,
34 enabling minority influence to occur (Clark & Maass, 1988; Moskowitz, 1996). In this way,
35 beliefs that women are competent and capable in the relevant work domain also facilitate
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 minority influence mechanisms because such perceptions are foundational for increased advice
4 taking and interpersonal influence (Mayer, Davis, & Schoorman, 1995; Nebus, 2006).
5
6

7
8 Moreover, we propose that the beliefs of the *team leader* are most essential because of
9
10 their formal position in the hierarchy and their control over whose voice is enacted in the course
11 of task accomplishment (Hollenbeck et al., 1995; Zaccaro, Rittman, & Marks, 2001). The highly
12 visible nature of their attitudes and actions also allows team leaders to disproportionately
13 influence the way the token female's voice is perceived in the team. We believe leaders' positive
14 gender beliefs may facilitate token female voice enactment in three ways. First, to the extent in
15 which leaders holding these beliefs exhibit more positive affective and evaluative reactions
16 toward a token female's suggestions (van Knippenberg et al., 2004), leaders offer reassurance to
17 the rest of the group that the token female is competent and attending to her voice will result in
18 valuable gains for enhancing task performance. Indeed, research has shown that a high-status
19 individual's endorsement can help legitimize the contributions of individuals who are relatively
20 unknown or foreign entities (Lin, 1999). This may reduce the male majority's tendency to
21 negatively stereotype the token female and marginalize her voice (Westphal & Milton, 2000) or
22 assumed lower competency (Ridgeway & Correll, 2004).
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 Second, leaders who react favorably to the token female's voice signal to the male
41 majority that she – like them – is a contributing member of the team. That is, members who see
42 their leaders responding in more positive and inclusive ways toward the token female may be
43 less likely to categorize her as an outgroup member and instead see her as a fellow team member
44 who shares the common objective of performing well on the team task. Indeed, research suggests
45 that when similarities are highlighted between minority and majority group members, minority
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 individuals can avoid or minimize outgroup biases and be evaluated more positively (Moscovici,
4
5 1985; Volpato, Maass, Mucchi-Faina, & Vitti, 1990).

6
7
8 Finally, leaders who ascribe to more positive gender beliefs can facilitate token female
9
10 voice enactment by protecting her and the rest of the group from social pressures to conform to
11
12 the majority. Indeed, majority members may find it interpersonally risky to adopt the minority
13
14 position (Moscovici, 1980), as doing so publicly presents a break from conformity with the
15
16 majority. The presence of a leader who not only reduces this social pressure but also encourages
17
18 the group to consider opposing perspectives is critical in ensuring that the minority is heard and
19
20 respected (Maier & Solem, 1952; Mannix & Neale, 2005).

21
22
23 In contrast, leaders who hold less favorable beliefs about women and their competence
24
25 and capabilities in a given domain are not only likely to legitimize and strengthen negative
26
27 evaluations that reduce the likelihood of token female voice enactment in male-dominated teams,
28
29 but also minimize the team's willingness to systematically process and be influenced by her
30
31 minority perspective. Thus, we expect that:

32
33
34
35 *Hypothesis 2: Leaders' gender beliefs moderate the effect of gender on voice enactment,*
36
37 *such that female tokens are more likely to have higher voice enactment than a*
38
39 *comparison male when leaders' gender beliefs are high than when they are low.*

40 41 42 **Task Complexity as a Boundary Condition of When Token Female Voice Enactment** 43 **Enhances Team Performance**

44
45 Assuming sufficient levels of token female voice enactment, there are several reasons
46
47 why her voice may have differential implications for team performance if enacted. First, the
48
49 token female's voice likely differs in content from that of the male majority and thus contributes
50
51 to the informational diversity in the team. Scholars have long argued that due to social role
52
53 differences in prior experiences, socialization, and societal expectations (Eagly & Wood, 1991;
54
55
56
57
58
59
60

1
2
3 Feingold, 1994), men and women often come to develop different skills, attitudes, and
4
5 approaches to task accomplishment. Men, for example, tend to be more agentic, focusing on
6
7 execution and attacking problems quickly and assertively (Eagly, 1987), whereas women tend to
8
9 be more interpersonally facilitative, collaborative, and democratic in their approach to problem-
10
11 solving (Eagly & Johannesen-Schmidt, 2001). To the extent these differences in how men and
12
13 women approach and process tasks also result in differing content of suggestions, we can expect
14
15 that, compared to enacting voice of the male majority, enacting token female voice will have
16
17 differential implications for team performance.
18
19

20
21 Second, there may be differences in the way token female voice is cognitively processed.
22
23 Research suggests that the perceived dissimilarity of a minority opinion can stimulate greater
24
25 systematic processing of the opinion and produce deeper understanding of the issue (De Dreu &
26
27 De Vries, 2001; Moscovici et al., 1994). This is even more likely when minority opinions are
28
29 expressed by categorically dissimilar individuals because the majority *expects* these individuals
30
31 to possess different information and ideas (Phillips, 2003; Phillips & Loyd, 2006). Thus, we
32
33 expect that this systematic processing and enactment of the token female's voice may produce a
34
35 fundamentally different understanding of the task that breaks uniformity, undermines the
36
37 dominant model, and introduces uncertainty by giving complexity to the situation (Moscovici &
38
39 Mugny, 1983). Indeed, work conducted by Nemeth and colleagues found that compared to
40
41 groups exposed to majority views, those exposed to minority views utilized more strategies for
42
43 problem solving (Nemeth & Kwan, 1985), engaged in broader information search (Nemeth &
44
45 Rogers, 1996), exhibited more originality (Nemeth & Kwan, 1987), and detected solutions that
46
47 otherwise would have gone undetected (Nemeth, 1986; Nemeth & Wachtler, 1983).
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 negatively impact team performance when task complexity is low. We do not expect this same
4
5 pattern of results for enacting the voice of a comparison male.
6

7
8 *Hypothesis 3: The relationship between token female voice enactment and team*
9
10 *performance is contingent upon task complexity such that this relationship is positive*
11
12 *when task complexity is high and negative when task complexity is low.*
13

14 15 **METHODS**

16 17 **Sample and Procedures**

18 Our sample consisted of 173 enlisted personnel (private to corporal, with 0-6 years of
19
20 service) in the United States Marine Corps. To achieve a causal understanding of how token
21
22 women newly integrated into an all-male team might fare with respect to voice enactment, male
23
24 participants were recruited from all-male direct-combat specialties within this branch of the
25
26 military, and female participants were recruited from combat support specialties that were
27
28 already gender integrated, such as administration, logistics, and motor transport. Participants
29
30 were recruited via flyers for a team decision-making study and offered gift cards totaling \$20
31
32 USD as compensation. There was no indication from the recruitment materials or procedure that
33
34 the study was about gender or voice. In our final sample, 8.2% of our participants were female,
35
36 which is comparable to the statistic of 6.8% active duty women serving in this branch of the
37
38 military (Statistic Brain, 2017). Participants' average age was 21.3 years, 61.2% self-identified
39
40 as White, 17.7% as Hispanic, 8.8% as Black, and 12.3% indicated "Other" as their ethnic group.
41
42
43
44

45 We conducted our study in a controlled training facility on a military base. Prior to
46
47 arriving at the research site, participants completed an online survey that measured demographics
48
49 and beliefs about women in the military. Upon arrival, participants were randomly assigned to 39
50
51 all-male or female token teams of four or five members, which reflected the standard size of
52
53 combat teams in this branch of the military. Twenty-two teams (56.4%) were composed of four
54
55
56
57
58
59
60

1
2
3 members, and 12 (30.1%) contained a token female. Participants were dressed in identical
4 uniforms, asked to remove all rank insignia prior to their arrival to de-emphasize existing
5 hierarchies, and randomly assigned a color-marked helmet for the purposes of identification –
6 blue, green, red, or yellow, and black in the case of five-person teams. An exception to this was
7 if a team contained a token female, who was always assigned the yellow helmet.
8
9

10
11
12
13
14 Each team attempted four tasks of varying task complexity, producing a within-subjects
15 design that generated 156 team observations in which each team served as its own control
16 comparison across the different tasks. Forty-eight observations involved female token teams, and
17 108 observations involved all-male teams. The four tasks were adapted from standardized tasks
18 used to train and evaluate officer candidacy in this branch of the military. Because the content of
19 the tasks themselves were high fidelity simulations of what participants would likely encounter
20 when deployed in combat, all tasks contained a physical component. The order of tasks
21 attempted by each team was counter-balanced and randomly assigned. Additionally, to mimic the
22 hierarchical structure of combat units, each member was randomly appointed as the “formal
23 leader” of one of the four tasks (in five-person teams, one team member did not have the
24 opportunity to lead).
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 **Description of Tasks**

41
42 The current consensus in the literature is that tasks high in complexity do not have
43 obvious solutions and therefore require innovative approaches (De Dreu & Weingart, 2003;
44 Morgeson & Humphrey, 2006). Thus, in our study, a task was considered high on *task*
45 *complexity* if the solution to the scenario was not readily apparent. Tasks low on task complexity
46 had relatively straightforward solutions. We describe each of our four tasks in detail below.
47
48
49
50
51
52

53 “*Casualty Rescue.*” (Appendix Task A.) The objective of this task was to cross a
54 “booby-trapped” room, climb through a 32-inch wide hole in the wall three feet off the ground,
55
56
57
58
59
60

1
2
3 locate and strap a 90-pound dummy to a stretcher, and move the dummy and stretcher back
4
5 through the hole to the starting point – all while avoiding off-limit areas. This task was low on
6
7 complexity because the solution – to use the provided 8-foot long wooden plank as a bridge to
8
9 safely cross into the first room, an up-ramp to access the hole in the wall, and a down-ramp from
10
11 the hole in the wall into the room with the dummy – was relatively straightforward.
12
13

14 **“Building Resupply.”** (Appendix Task B.) The objective of this task was to transport five
15
16 17-pound ammunition containers and all members of the team into a building through a window
17
18 that was eight feet off the ground – all while avoiding off-limit areas. The only way into the
19
20 building was to scale a set of parallel benches and a stone pedestal measuring approximately six
21
22 feet in height. Equipped with a 10-foot long plank, successful teams passed the plank back and
23
24 forth between the two segments of the task until all members of the team and the five
25
26 ammunition containers were in the building. This task was low in complexity because the
27
28 solution was straightforward.
29
30
31

32 **“Medicine Delivery.”** (Appendix Task C.) The objective of this task was to transfer a 5-
33
34 pound container from one side of a large room to the other side without throwing it or stepping
35
36 into a large, off-limit space separating the two sides. Teams were split such that half the team
37
38 started on one side of the room with the container and the other half was on the opposite side to
39
40 receive the container. A variety of materials were scattered in adjacent rooms, including some
41
42 distractor items, such that teams had to search and identify relevant and usable items. Key
43
44 resources were a 40- and 18-foot long rope, a 6-foot long plank, and a hook. The most common
45
46 solution implemented by teams was to combine the two lengths of rope to devise a pulley system
47
48 or a zip line via the ceiling rafters to gently transfer the container across the divide. This task was
49
50 high in complexity because the solution was not readily obvious, and the presence of distractors
51
52
53
54
55
56
57
58
59
60

1
2
3 required the team to deliberate on which resources were essential and how to fashion those
4
5 resources into a workable solution.
6

7
8 **“Chemical Walls.”** (Appendix Task D.) The objective of this task was to transfer five 40-
9
10 pound water jugs and all team members over a 10-foot high wall with two perpendicular sides.
11
12 All surfaces of the wall and areas of the ground near the wall were marked as off-limits. Only the
13
14 top of the wall was safe to touch. Teams were given a variety of materials to use, including some
15
16 distractor items, requiring teams to identify relevant and usable items. Key resources were a 6-
17
18 and 8-foot long plank. The most common solution implemented by teams was to lay both planks
19
20 on top of the L-shaped corner of the wall to create a platform onto which members could safely
21
22 lift themselves and the water jugs. This task was high in task complexity because the solution
23
24 was not obvious, and the presence of distractors required the team to deliberate on which
25
26 resources to use and how they could produce a workable solution.
27
28
29

30
31 **Manipulation checks for task complexity.** We consulted 17 doctoral students in a
32
33 research-intensive management department as subject matter experts (SMEs). Given a definition
34
35 of task complexity and details about each task (including the initial set up, materials provided,
36
37 and most common solution), SMEs rated each task on three items adapted from Morgeson and
38
39 Humphrey (2006) about the extent the task required “generating unique ideas,” “generating
40
41 innovative solutions,” and “cognitive processing” ($\alpha = .91$). The SMEs rated the two high
42
43 complexity tasks higher than the two low complexity tasks ($M_{\text{Casualty Rescue}} = 2.90$, $M_{\text{Building Resupply}}$
44
45 $= 3.14$, $M_{\text{Medicine Delivery}} = 4.10$, and $M_{\text{Chemical Walls}} = 3.84$). Results also showed that the average
46
47 SME ratings of the two high complexity tasks ($M = 3.97$, $SD = .94$) differed significantly from
48
49 the SME ratings of the two low complexity tasks ($M = 3.02$, $SD = 1.16$; $F_{(1, 66)} = 13.78$, $p < .05$).
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Moreover, the SMEs did not differentiate between the two low complexity tasks nor the two high
4 complexity tasks. Thus, we coded each task as high or low in complexity accordingly.¹
5
6

7 8 **Measures**

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

10
11 **Gender Composition.** We coded each team as 0 = an all-male team, or 1 = a female
12 token team (i.e., one lone female with all-male teammates).
13
14

15
16 **Voice and Voice Enactment.** We assessed voice and voice enactment using observer
17 ratings from four hypothesis-blind doctoral students. Using the color-marked helmets to identify
18 participants, observers coded in real time who voiced a suggestion and whether the team acted on
19 it. Consistent with voice being defined as suggested actions intended to improve the team's task
20 accomplishment (Premeaux & Bedeian, 2003), observers coded voice as instances where
21 members proposed a new plan for accomplishing an element of the task. Suggestions were
22 strategic (e.g., "How about we try to get the rope to the other side and just sort of 'zip-line' it?")
23 or tactical (e.g., "Lift your end onto the plank and we'll see-saw it"). Examples of non-
24 suggestions included coordinating or directing behaviors that requested specific actions from
25 specific members, or assigning members to roles and subtasks (Bales, 1950; Lord, 1977).
26
27
28
29
30
31
32
33
34
35
36
37
38

39 Each task was assigned a primary observer who coded all voice and voice enactment that
40 occurred in that task. To establish the validity of the primary observer's ratings, we assessed
41 inter-rater agreement and reliability for voice enactment using 201 person episodes² (29% of our
42 total observations) in which a second observer simultaneously coded research participant
43
44
45
46
47
48
49

50 ¹ We also analyzed our data by substituting the binary coding to task complexity with the SME's average scores
51 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
52 $M_{\text{Chemical Walls}} = 3.84$). Results were consistent with that of our original analyses.

53 ² By person episode, we mean how much each individual's voice was enacted in a given task. With 173 participants
54 engaged in 4 tasks, the total number of person episodes observed was $173 \times 4 = 692$. Of these person episode
55 observations, 201 (i.e., 29%) were double-coded and thus used to assess inter-rater agreement and reliability.
56
57
58
59
60

1
2
3 behavior. Because voice enactment was a continuous count variable, we calculated intra-class
4 correlations to assess whether the ratings of primary observers were reliable. Used in this way,
5
6 higher ICC values result when the correlation between two sets of ratings is high and the
7
8 variance between observer ratings across targets is low (i.e., higher inter-rater agreement and
9
10 reliability; Shrout & Fleiss, 1979). Results from a one-way ANOVA confirmed that the degree of
11
12 agreement across observers was acceptable ($ICC_1 = .43$ and $ICC_2 = .60$ ($F_{(200,201)} = 2.50, p < .05$;
13
14 Bliese, 2000). Thus, only the primary observers' ratings were used for data analyses.
15
16
17
18

19 To ensure that voiced suggestions represented a change to the status quo, we only
20 analyzed suggestions made after the very first suggested action as voice. We did this because
21
22 research has shown that the first suggestion often anchors the team's discussion and approach to
23
24 the task (Berg, 2014) and represents the initial status quo that subsequent voice aims to change
25
26 and improve upon (Premeaux & Bedeian, 2003). Thus, voice frequency reflected the total
27
28 number of times a member spoke up with a suggestion after the first suggestion was made, and
29
30 voice enactment reflected the total number of these suggestions that were enacted. In four-person
31
32 teams, we accounted for the missing values for a hypothetical fifth member's voice in our
33
34 analyses by mean replacement (i.e., averaged the voice of blue, green, and red helmet members).
35
36
37
38
39

40 **Leader Gender Beliefs.** We assessed participants' beliefs about the capabilities of
41
42 women in the military in the pre-survey prior to team composition or leader assignment with
43
44 three representative items from a pool of competence-related items outlined in Boldry, Wood,
45
46 and Kashy's (2001) study of gender stereotypes in a military context. These items focused on
47
48 three attributes – physical fitness, leadership, and readiness to engage in military operations –
49
50 that reflected the biggest objections to women in the military (Enloe, 1983; Holland, 2006). We
51
52 expected that to the extent to which participants held beliefs contrary to these objections, they
53
54
55
56
57
58
59
60

1
2
3 would hold more positive expectations of the contributions of the token female and be more
4 likely to enact her voice. Items were measured on a 7-point scale ($1 = \textit{strongly disagree}$ to $7 =$
5 *strongly agree*) using the stem “In your opinion, to what extent are women in the military ...”
6 followed by three items “physically strong”, “combat ready”, and “capable leaders” ($\alpha = .90$).
7
8
9

10
11
12 **Team Performance.** The primary observer for the task evaluated each team’s
13 performance in terms of completion time, which was measured via stopwatch and recorded as
14 soon as the team completed the task. Although speed (efficiency) and accuracy (effectiveness)
15 are often depicted as tradeoffs (Fitts, 1954; Woodworth, 1899), completion time in our study
16 captured both elements of performance. Across all tasks, the only way to achieve a faster
17 completion time was to perform the task more accurately by enacting successful strategies that
18 advanced the task in measurable ways. In contrast, pursuing low accuracy strategies or safety
19 violations that incurred a 60-second time penalty resulted in a longer completion time.
20
21
22
23
24
25
26
27
28
29

30
31 Consistent with how these tasks are typically administered, each team was allocated 18
32 minutes per task. Teams that did not complete the task within the 18 minutes were assigned the
33 maximum time of 18, indicating the lowest performance. In a post-hoc examination of cases
34 where teams did not complete the task, we found no case where, if given a reasonable amount of
35 additional time, the team would have completed the task. There were also no teams that
36 completed the task in exactly 18 minutes. In order to create a performance index where higher
37 scores reflected better performance, we subtracted the team’s completion time from 18. Thus,
38 our performance index ranged from 0 (i.e., those who did not complete the task) to 13.32 (i.e.,
39 those who completed the task in 4.68 minutes, the fastest time recorded).
40
41
42
43
44
45
46
47
48
49
50

51
52 **Control Variables.** We included several control variables that potentially could impact
53 interpersonal processes in the team and team performance aside from gender composition. Using
54
55
56
57
58
59
60

1
2
3 demographic data captured prior to team life, we controlled for diversity in rank and ethnicity
4
5 using Blau's index of heterogeneity (1977) and age diversity using the standard deviation
6
7 (Harrison & Klein, 2007). We controlled for team size, as larger teams might perform better.
8
9 Because of fatigue effects and the potential effects of accrued experience on team interactions
10
11 (Finseraas et al., 2016), we controlled for task order, which indicated whether the episode in
12
13 question was the first, second, third, or fourth task attempted. We also controlled for the physical
14
15 demand of the task. Following the Dictionary of Occupational Titles' (1977) definition of
16
17 physical demands as the estimation of the overall strength requirement of the task, we divided
18
19 the total weight of objects and human bodies to be moved from point A to point B by the number
20
21 of team members and used the average weight per team member to categorize each task as low
22
23 (≤ 20 lbs. of force) or high (≥ 100 lbs. of force) physical demand. Consequently, we coded
24
25 Casualty Rescue and Medicine Delivery as 0 (low physicality) and Building Resupply and
26
27 Chemical Walls as 1 (high physicality). Finally, we controlled for voice frequency to isolate
28
29 variance explained in voice enactment beyond the amount of voice that was expressed.³
30
31
32
33
34

35 **Data Analyses.** Given the multilevel nature of our data (task episodes nested within
36
37 teams) and evidence that team membership accounted for significant variance in team
38
39 performance ($ICC_1 = .11$, $F_{(38, 116)} = 1.50$, $p < .05$), we used hierarchical linear modeling for our
40
41 analyses (HLM 6.2; Raudenbush, Bryk, Cheong, Congdon, & du Tiot, 2004). Each task episode
42
43 and its associated amount of voice enactment, task complexity, and team performance were
44
45 treated as Level 1 variables. Because the person in the leader position was rotated for each task
46
47 episode, leader's gender beliefs and other episodic control variables were also treated at Level 1.
48
49 Gender composition of the team and team size were treated as Level 2 variables.
50
51
52

53
54
55 ³ For all conclusions relevant to hypothesis testing, the omission of these controls did not change results.
56
57
58
59
60

1
2
3 To isolate the pure effects of the Level 1 variables of interest on outcome variables within
4 a task episode, we group mean centered task complexity, voice frequency, and voice enactment
5 (Hofmann, & Gavin, 1998; Enders & Tofighi, 2007). This allowed Level 1 variables to be
6 expressed relative to the other cases within the same level of analysis, which in our case, is the
7 task episode. For example, voice enactment is considered with respect to that of other team
8 members' voice enactment in the same episode, yielding a slope coefficient that is unaffected by
9 between group differences other than the Level 2 variables under study. Group mean centering
10 Level 1 variables is also recommended when cross-level interactions are of interest, which was
11 the case as we sought to compare Level 1 relationships in all-male versus female token teams
12 (Bauer & Curan, 2005). Following the recommendations of Hofmann and Gavin (1998), we
13 reintroduced the mean level of voice enactment at Level 2 to ensure that our cross-level
14 interactions were not spurious. To assess model fit, we used a deviance index reported in HLM
15 analysis (Bryk & Raudenbush, 1992), where smaller deviance values indicate a better model fit.
16 These indices are reported in Tables 4 and 5.

17
18
19 **Analyses Involving a Comparison Male.** Because token females only resided in female
20 token teams, we used the voice enactment of a randomly assigned male counterpart from an all-
21 male team as a comparison score as opposed to a randomly selected male from a female token
22 team. Given that simply being exposed to minority perspectives can increase divergent thought
23 and processing in majority members (Nemeth, 1986; Phillips et al., 2004; Nemeth & Nemeth-
24 Brown, 2003), it is possible that input from the token female may fundamentally change the
25 content of how majority male members think and what they choose to voice. These spillover
26 effects, in turn, may potentially produce greater receptivity to token female voice or, in certain
27 tasks, increase male voice enactment at the expense of enacting her voice. In either case, a
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 comparison of token female voice enactment to male voice enactment cannot be cleanly
4
5 observed when both speakers are on the same team.
6

7
8 Thus, to create a cleaner comparison, in Hypotheses 1 and 2, we compared the voice
9
10 enactment of the token woman – who always wore a yellow helmet – to that of the male who
11
12 was randomly assigned the yellow helmet in an all-male team. Moreover, when examining the
13
14 effects of token female voice enactment on team outcomes (Hypothesis 3), we examined the
15
16 cross-level interaction between gender composition, yellow helmet voice enactment, and task
17
18 complexity on team performance, while controlling for the voice frequency and enactment of all
19
20 other helmet colors (blue, green, red, and black) as well as their respective interactions with team
21
22 gender composition and task complexity. This allowed us to assess the unique contribution of the
23
24 token female’s voice enactment on team performance, beyond the effects of others’ voices.
25
26
27

28 RESULTS

29
30 Means, standard deviations, and correlations among study variables are shown in Table 1.
31
32 Average voice frequency ($F_{(1, 154)} = .43, n.s.$) and voice enactment ($F_{(1, 154)} = .00, n.s.$) in female
33
34 token versus all-male teams were not significantly different. However, token females engaged in
35
36 more voice behavior ($M_{\text{female}} = 5.54, SD_{\text{female}} = 4.09$) compared to yellow-helmeted males in all
37
38 male teams ($M_{\text{male}} = 3.51, SD_{\text{male}} = 3.19; F_{(1,155)} = 11.28, p < .05$).
39
40
41

42 -----
43 Insert Table 1 about here.
44 -----
45

46 Hypotheses 1a and 1b examined whether token female voice enactment in female token
47
48 teams was lower or higher compared to that of a comparison male. To test Hypotheses 1a and 1b,
49
50 we examined the direct effects of gender composition on yellow helmet voice enactment. Results
51
52 showed that gender composition positively predicted yellow helmet voice enactment ($\gamma = .93, p <$
53
54 $.05$; Model 2 of Table 2), indicating that even after controlling for voice frequency, token female
55
56
57
58
59
60

1
2
3 voice enactment in female token teams was higher than a comparison male's voice enactment in
4 an all-male team.⁴ Thus, Hypothesis 1b (and not 1a) was supported.

7
8 Hypothesis 2 predicted leaders' gender beliefs would moderate the effect of gender on
9 voice enactment, such that female tokens are likely to have higher voice enactment than a
10 comparison male when leaders' gender beliefs are high than when they are low. As shown in
11 Model 4 of Table 2, results indicated that gender composition interacted with leader's gender
12 beliefs to predict yellow helmet voice enactment ($\gamma = .56, p < .05$). Shown in Figure 1, leader
13 gender beliefs was positively related to token female voice enactment in female token teams ($\gamma =$
14 $.50, p < .05$) but unrelated to the yellow-helmeted male's voice enactment in all-male teams ($\gamma =$
15 $-.02, n.s.$). This suggests that the presence of positive leader gender beliefs was essential for high
16 token female voice enactment, but, interestingly, there appeared to be no penalty on token female
17 voice enactment at low levels of leader gender beliefs. Moreover, these results were robust even
18 when we removed cases where the token female was in the leadership role, suggesting that the
19 effects of leader gender beliefs hold for male and female leaders. Lastly, leader gender beliefs
20 had no effect on the yellow-helmeted male's voice enactment, suggesting that gender beliefs
21 were uniquely activated in the presence of a token female, offering support for the validity of our
22 gender beliefs measure in this context. Thus, Hypothesis 2 was supported.

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Insert Table 2 and Figure 1 about here.

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.

1
2
3 Hypothesis 3 predicted the relationship between token female voice enactment and team
4 performance is contingent upon task complexity such that this relationship is positive when task
5 performance is contingent upon task complexity such that this relationship is positive when task
6 complexity is high and negative when task complexity is low. As shown in Model 4 of Table 3,
7 the three-way interaction term of team gender composition x yellow helmet voice enactment x
8 task complexity was significant ($\gamma = 1.21, p < .05$). For ease of interpretation, in Figure 2, we
9 split the interaction plot between yellow helmet voice enactment and task complexity into two
10 graphs corresponding to female token teams versus all-male teams. Simple slopes analyses
11 indicated that in female token teams, the relationship between token female voice enactment and
12 team performance was positive and significant when task complexity was high ($\gamma = .50, p < .05$)
13 but negative when task complexity was low ($\gamma = -.45, p < .05$). Additionally, results showed that
14 in all-male teams, yellow helmet voice enactment was unrelated to team performance when task
15 complexity was high ($\gamma = -.20, n.s$) or low ($\gamma = -.03, n.s$). Thus, our results provided support for
16 Hypothesis 3. Notably, these relationships did not hold when we replaced token female voice
17 enactment with voice frequency, suggesting that when linking her contributions to team
18 outcomes, it is more important to focus on enacted voice as opposed to simply speaking up.⁵
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 -----
39 Insert Table 3 and Figure 2 about here.
40 -----

41 **Supplementary Analyses**

42 Although the above results provided support for our hypothesized model, we conducted
43 several supplementary analyses to further understand the micro-dynamics of voice in diverse
44 teams in our unique context (Hollenbeck & Wright, 2017).
45
46
47
48
49
50
51

52 ⁵ Following Aguinis, Gottfredson, and Joo (2013), we identified outliers in token female voice and voice enactment
53 as +2.24 SD units or more above the mean within each task. 2 outliers pertained to voice enactment, and 9 outliers
54 pertained to voice frequency. The results for Table 2 and Table 3 remained the same, despite excluding these outlier
55 observations from the analyses.
56
57
58
59
60

1
2
3 **The Impact of Voice Quality.** One competing explanation for why the token female's
4 voice was enacted is because her suggestions were expected to be or were of higher quality, and
5 therefore facilitated team performance in complex tasks. We did not measure voice content or
6 indicators of voice quality, but we tested this possibility by using survey data from female token
7 teams that captured members' pre-task expectations prior to the team's attempts on any task ("To
8 what degree do you expect this individual to offer valuable contributions helpful to the team's
9 success?" $1 = \text{not at all}$ to $7 = \text{a very large extent}$) and post-task perceptions of the quality of
10 contributions for each of the four tasks ("During the task today, to what degree did this
11 individual offer valuable contributions helpful to the team's success?" $1 = \text{not at all}$ to $7 = \text{a}$
12 *very large extent*). Although pre-task evaluations of the token female's expected contributions
13 ($M_{\text{female}} = 4.68$, $SD = 1.39$) was numerically lower than that of the yellow-helmeted male in all-
14 male teams ($M_{\text{males}} = 4.87$, $SD = 1.42$), this difference was not significant ($F_{(1, 132)} = .49$, *n.s.*).
15 Similarly, the post-task evaluations of the token female ($M_{\text{female}} = 5.37$, $SD = 1.53$) was
16 numerically lower than that of the yellow-helmeted male in all-male teams ($M_{\text{males}} = 5.53$, $SD =$
17 1.41), but this difference was not significant ($F_{(1, 531)} = 1.45$, *n.s.*).
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 **The Impact of Task Complexity.** Another competing explanation is that the token
39 female's voice was enacted more in complex tasks because more difficult and uncertain tasks
40 increase majority members' openness to minority proposals, whereas in less complex tasks,
41 majority members perceive minority input as creating unnecessary division. To explore this
42 possibility, we examined whether task complexity interacted with the gender composition of the
43 team to predict voice enactment. Regardless of gender composition, there was less yellow helmet
44 voice enactment in high complexity tasks than in low complexity tasks ($\gamma = -.82$, $p < .05$), likely
45 because it was harder to discern which suggestions should be enacted in highly complex tasks.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ANOVA comparisons further revealed that token female voice enactment ($M_{\text{female}} = 3.65$) was
4 higher than that of yellow helmeted-males ($M_{\text{male}} = 1.65$, $F_{(1,75)} = 9.69$, $p < .05$) in low
5 complexity tasks, but her voice enactment ($M_{\text{female}} = 1.33$) was not significantly different from
6 yellow-helmeted males ($M_{\text{male}} = 1.46$, $F_{(1,75)} = .07$, $n.s.$) in high complexity tasks. Ironically, this
7 suggests that female voice was enacted more frequently in low complexity tasks (when enacting
8 her voice appeared to harm team performance), but not more likely enacted in high complexity
9 tasks (when enacting her voice helped performance).
10
11
12
13
14
15
16
17
18

19 **The Impact of Leader's Gender Beliefs versus Team Beliefs.** Although our hypothesis
20 and results highlighted the impact of leader's gender beliefs on token female voice enactment in
21 female token teams, it is also possible that fellow teammates' positive expectancies about
22 women's competence and capabilities (or even the token female's own positive beliefs) will
23 facilitate her voice enactment. We amended Model 3 of Table 1 to include two interaction terms
24 – one between non-yellow helmet beliefs and gender composition to reflect fellow male team
25 members' beliefs, and one between gender composition and yellow helmet beliefs to reflect the
26 token female member's beliefs. Neither interaction term significantly predicted yellow helmet
27 voice enactment ($\gamma = .07$, $n.s.$ and $\gamma = .29$, $n.s.$, respectively), indicating that above and beyond
28 the leader's beliefs, neither the gender beliefs of male team members nor that of the token female
29 explained additional variance in token female voice enactment in female token teams.
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 **The Impact of Gender Composition on Deliberation.** According to the minority
45 influence literature, one explanation for the positive impact of token female voice enactment on
46 complex tasks (and its negative impact on less complex tasks) is that minority voice produces
47 more divergent and systematic processing in the team, which in turn slows down the information
48 processing and decision-making of the team prior to taking action. To explore this possibility, we
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 examined whether female token teams engaged in more voice before enacting a suggestion
4 compared to all-male teams. Results showed that more suggestions occurred before the first
5 enacted suggestion in female token teams ($M = 3.34$, $SD = 3.70$) compared to all-male teams (M
6 $= 1.63$, $SD = 2.99$; $F_{(1, 84)} = 5.36$, $p < .05$). This corroborates theoretical arguments that exposure
7 to minority viewpoints stimulates a wider search for information and the detection of multiple
8 solutions (Nemeth & Rogers, 1996), and suggests that the lack of a minority viewpoint –
9 especially in all-male teams because of their agentic nature – produced a proclivity toward action
10 over more deliberative strategies. It seems, then, that in our context, the token female provided a
11 “check and balance” on the tendency of all-male teams to engage in action-first strategies.
12
13
14
15
16
17
18
19
20
21
22
23

24 **The Impact of Majority Member Voice.** One might also conjecture about the role of
25 male-majority (i.e., non-yellow helmet) voice on team performance in more or less complex
26 tasks. The results of this analysis detected a significant three-way interaction between non-
27 yellow helmet voice enactment, gender composition, and task complexity ($\gamma = -4.19$, $p < .05$; see
28 Figure 3). Simple slopes analyses revealed that in female token teams, male majority voice
29 enactment was *negatively* related to team performance when the task was complex ($\gamma = -.71$, $p <$
30 $.05$) but positively related to team performance when the task was low in complexity ($\gamma = 2.05$, p
31 $< .05$). This finding is the mirror opposite of the pattern we see in Figure 2 concerning token
32 female voice enactment and consistent with research suggesting that majorities induced better
33 performance in tasks where convergent thinking was useful and the perspective of the majority
34 was appropriate (Nemeth, Mosier, & Chiles, 1992). Taken together, these results suggest that in
35 female token teams, there may be something uniquely beneficial about enacting the token
36 female’s voice and something uniquely harmful about enacting the male majority’s voice in high
37 complexity tasks while the opposite holds true for low complexity tasks.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 (Fiske, Cuddy, Glick, & Xu, 2002; Mannix & Neale, 2005) and also highlight more broadly the
4 importance of factors outside of women to facilitate female voice enactment and help move the
5 needle toward gender parity.
6
7
8
9

10 Our third contribution is to clarify when enacting the token woman's voice benefits team
11 performance. Our findings are consistent with past research suggesting that diverse teams can
12 outperform homogeneous teams on complex tasks (van Knippenberg et al., 2004) and that
13 greater female representation is beneficial for solving problems, including ones involving
14 realistic military scenarios (LePine, Hollenbeck, Ilgen, Colquitt, & Ellis, 2002; Hirschfeld,
15 Jordan, Field, Giles, & Armenakis, 2005). Still, we extend this work by showing that it was
16 token female voice enactment specifically that enhanced performance in cognitively demanding
17 tasks. Identifying voice enactment and *whose* voice is enacted as a novel mechanism offers a
18 new perspective on the processes underlying the effects of diversity on performance. In doing so,
19 we address calls to investigate the micro-dynamics of diverse teams and elucidate *how* the
20 behavior of certain members and the reactions they elicit can exert disproportionate effects on
21 team outcomes for both good and ill (Emich & Wright, 2016; van Knippenberg & Mell, 2016).
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 Finally, we contribute to minority influence theory. Although some work has examined
39 the efficacy of double minorities in influencing the way dominant members think about
40 minority-relevant social issues (Kelley, 1971), it is unclear whether these effects generalize
41 beyond experimental settings to affect the processes and performance of teams engaged in
42 physically and cognitively demanding tasks (Maass & Clark, 1984). Relatedly, whereas the
43 minority influence literature has focused on presentation tactics the minority individual should
44 employ to increase his/her influence (Moscovici, 1980), our finding that leader gender beliefs
45 and task complexity act as facilitators of both token female voice enactment and performance
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 provides a broader understanding of factors outside of the minority individual that shape their
4
5 impact on real team tasks. Taken together, our work demonstrates how minority influence theory
6
7 can be extended and applied to address gender-based voice inequities in organizations.
8
9

10 **Practical Implications**

11 Our findings indicate that members and managers of female token teams need to pay
12
13 attention to whose voice is enacted, because enacting the token female's (versus the male-
14
15 majority's) voice exerts unique benefits (or costs) for team performance. If the task is high in
16
17 complexity, token females should speak up with unique ideas, and managers should seek to
18
19 enhance the token female's voice enactment over that of the male majority because it may
20
21 encourage the team to generate innovative performance strategies required to accomplish such
22
23 tasks successfully. However, if the task is low in complexity and requires speedy completion,
24
25 token females should focus on speaking up in ways that reflect the majority approach to
26
27 efficiently reach a workable solution, and managers should be cautioned against enacting
28
29 divergent voice as it may slow down the work.
30
31
32
33

34 Moreover, managers should recognize that merely having a token female present does not
35
36 distinguish her team's performance from that of all-male teams. Rather, this woman has to speak
37
38 up and, more importantly, her suggestions have to be enacted. A long history of gender diversity
39
40 research would suggest that, due to tokenism, stereotype threat, and gender role expectations of
41
42 women, female members may not speak up – and indeed, our field notes indicated that some of
43
44 our female participants were passive, remained silent, or simply followed the directions of the
45
46 other members. Yet, as our results show, when the token female did speak up, minority influence
47
48 mechanisms appeared to work in favor of her voice enactment, and our findings suggest one way
49
50 to further facilitate token female voice enactment is to select or train team leaders to hold more
51
52 positive beliefs about women's capabilities in the relevant task domain.
53
54
55
56
57
58
59
60

1
2
3 As for practical implications in the military context specifically, those opposed to the
4 integration of women focused much of their arguments on physical differences between the
5 genders. Indeed, compared to all-male teams, female token teams performed worse on high
6 physicality tasks, and our field notes indicated multiple instances where the token female was
7 left behind in a physically demanding task because she could not independently execute a
8 physical element of the task. Even so, the poorer performance exhibited by all-male teams on
9 high complexity tasks suggests that we need to look beyond a purely physical perspective to
10 truly understand the pitfalls and opportunities of gender integration in the military. If recent
11 history is any indication, many military mistakes that led to unnecessary loss of life or capture
12 were more likely to result from poor tactical decision-making than from a lack of physical
13 ability. Thus, the conversation regarding the liabilities of females in this context when it comes
14 to physically demanding tasks should, at the very least, be complemented by a conversation of
15 the severe liabilities of all-male teams in making tactical decisions on complex tasks.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32 **Limitations and Directions for Future Research**

33 We believe there are several noteworthy methodological strengths of this study. By
34 manipulating gender composition and task complexity, randomly assigning individuals to teams,
35 and rotating individuals through leadership roles, our within-subjects study design eliminated
36 many potential alternative explanations that might obscure the effects of voice enactment on
37 team performance. Moreover, our use of trained, reliable, and hypothesis-blind observers
38 reduced biases that might be associated with retrospective self-reports, and observing all teams
39 on all four tasks in a counter-balanced order mitigated confounds that might be attributable to
40 fatigue or different stages of team development. Lastly, we examined representative individuals
41 on representative tasks in a context in which gender-integration is a salient and contemporary
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 issue. Taken together, these methodological strengths help to mitigate endogeneity concerns of
4 self-selection and increase the internal and external validity of our findings.
5
6

7
8 Nevertheless, there are several key limitations in our work. Our sample contained very
9
10 few women – twelve in all – who were recruited from a gender-integrated function to be token
11 females in our teams. Although the small sample and functional background of these women was
12 unavoidable given the context of our study, we are unable to rule out the possibility that there
13 was something particular about these women that may have advantaged their voice enactment in
14 comparison to their male counterparts. Our supplementary analyses did not reveal systematic
15 differences between the men and women in our sample with regards to expected or perceived
16 quality of contributions, but it is possible that these women possessed higher competence
17 (Rosette & Tost, 2010), cognitive ability, or specialized skills and training in their occupational
18 function that contributed to their likelihood of voice enactment. Future work seeking to replicate
19 our findings, thus, may increase the number of female participants, as well as control for
20 individual differences in ability. As a related limitation, because our female and male
21 participants were drawn from different departments, we acknowledge that our results may be in
22 part driven by functional diversity rather than gender diversity. We think this is unlikely, because
23 if our results were indeed driven by functional diversity, the moderating effects of leaders’
24 gender beliefs would not have been expected. Still, we cannot completely rule out the possibility
25 of functional diversity as an alternative explanation for our findings. Future work may improve
26 upon ours by controlling for function when comparing all-male versus female token teams.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 There is more work to be done on gender and voice enactment. Because this is the first
50 study to examine voice enactment in teams, observing voice in real time during team task
51 accomplishment was necessary to ensure an accurate interpretation of the phenomenon.
52
53
54
55
56
57
58
59
60

1
2
3 However, this approach also prevented us from assessing gender or individual differences in *how*
4 the voice was presented that may impact the likelihood of enactment. For example, men may
5 present suggestions directly whereas women may present suggestions as questions (e.g., Maltz &
6 Borker, 1982). Likewise, the minority influence literature suggests that consistency, confidence,
7 and persistence in one's position can be important for influencing the majority (e.g., Moscovici,
8 Lage, & Naffrechoux, 1969). Our field notes highlighted the relevance of these factors in one
9 instance where the token female proposed the correct solution to a complex task. Even though
10 the rest of the team moved away from her suggestion, she persisted. Once the team eventually
11 enacted her suggestion, the team was successful. We encourage future research to develop
12 reliable instruments for assessing voice presentation. Future work may also track specific
13 instances of voice and capture the quality of those suggestions to yield a more precise
14 understanding of how enacting a particular suggestion facilitates team performance.

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31 There are several context-specific factors that may have influenced our findings. Basing
32 our four tasks on scenarios used by this institution to evaluate leadership and teamwork skills
33 improved the ecological validity of our study, but we acknowledge that our task complexity
34 manipulation may not have been as "controlled" as what one might encounter in a traditional
35 laboratory experiment. Although we feel that trading ecological validity for comparability still
36 yielded meaningful results, especially in view of our manipulation checks, we encourage future
37 research to replicate our findings by utilizing standardized high and low complexity tasks, tasks
38 with no physical component, or tasks where team outcomes are less time-sensitive. Moreover,
39 given that teams in the military are characterized by a relatively high degree of authority
40 differentiation (Hollenbeck, Beersma, & Schouten, 2012), the effect we documented for leaders'
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 gender beliefs in the emergence of female voice enactment may be especially strong. Future
4
5 research may investigate these relationships in contexts where authority differentiation is low.
6

7
8 Future research may also focus more on why leaders' gender beliefs matter. We theorized
9
10 that these beliefs facilitate token female voice enactment because of more positive evaluations of
11
12 the token female's voice, greater valuation of the uniqueness of her voice, or both, but we did not
13
14 capture any of these mechanisms. Relatedly, we did not capture leaders' beliefs about the men's
15
16 competence and capabilities, so it is unclear whether leaders' beliefs about women were as high
17
18 or higher than those about men and how those two sets of beliefs ultimately shape choices to
19
20 enact the token female's versus the male majority's voice. Future research may investigate these
21
22 possibilities and whether leader's gender beliefs produced specific behaviors, such as soliciting
23
24 voice from token females, that increased the likelihood of her ideas being enacted.
25
26
27

28
29 Finally, future research may examine the role of voice enactment in teams marked by
30
31 other forms of diversity. Given that a large number of occupations with a history of gender
32
33 segregation are increasingly confronted with the realities of integration, we believe our findings
34
35 on voice enactment extend existing research on gender diversity and team performance in
36
37 meaningful ways. However, as diversity comes in many different forms, research on the possible
38
39 effects of other types of diversity on voice enactment and team performance is also warranted. It
40
41 may be that, depending on the type of diversity, majority versus minority voice enactment have
42
43 different implications for team performance.
44
45

46 47 **Conclusion**

48
49 By integrating the employee voice, team diversity, and minority influence literatures, this
50
51 study elucidates the unique role of token female voice enactment in traditionally male-dominated
52
53 teams and identifies the conditions that facilitate its emergence and impact on team performance.
54
55 Our work seeks to stimulate future study of not only voice in diverse teams but *whose* voice is
56
57
58
59
60

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.

- 1
2
3 De Dreu, C. K., & West, M. A. 2001. Minority dissent and team innovation: The importance of
4 participation in decision making. *Journal of Applied Psychology*, 86: 1191-1201.
- 5 DeRue, D. S., Nahrgang, J. D., Wellman, N., & Humphrey, S. E. 2011. Trait and behavioral
6 theories of leadership: An integration and meta-analytic test of their relative
7 validity. *Personnel Psychology*, 64(1): 7-52.
- 8 Detert, J. R., & Burris, E. R. 2007. Leadership behavior and employee voice: Is the door really
9 open? *Academy of Management Journal*, 50: 869-884.
- 10 Deutsch, M., & Gerard, H. B. 1955. A study of normative and informational social influences
11 upon individual judgment. *The Journal of Abnormal and Social Psychology*, 51(3): 629-636.
- 12 Eagly, A. H. 1987. Reporting sex differences. *American Psychologist*, 42: 756-757.
- 13 Eagly, A. H., & Johannesen-Schmidt, M. C. 2001. The leadership styles of women and men.
14 *Journal of Social Issues*, 57: 781-797.
- 15 Eagly, A. H., & Wood, W. 1991. Explaining sex differences in social behavior: A meta-analytic
16 perspective. *Personality and Social Psychology Bulletin*, 17: 306-315
- 17 Edmondson, A. C. 2003. Speaking up in the operating room: How team leaders promote learning
18 in interdisciplinary action teams. *Journal of Management Studies*, 40(6): 1419-1452.
- 19 Edmondson, A. C. 2012. Teamwork on the Fly. *Harvard Business Review*, 90(4): 72-80.
- 20 Edmondson, A. C., Bohmer, R. M., & Pisano, G. P. 2001. Disrupted routines: Team learning and
21 new technology implementation in hospitals. *Administrative Science Quarterly*, 46(4): 685-
22 716.
- 23 Ely, R. J., & Thomas, D. A. 2001. Cultural diversity at work: The effects of diversity
24 perspectives on work group processes and outcomes. *Administrative Science Quarterly*,
25 46(2): 229-273.
- 26 Emich, K. J., & Wright, T. A. 2016. The 'I's in team: The importance of individual members to
27 team success. *Organizational Dynamics*, 45(1): 2-10.
- 28 Enders, C. K., & Tofighi, D. 2007. Centering predictor variables in cross-sectional multilevel
29 models: A new look at an old issue. *Psychological Methods*, 12(2): 121-138.
- 30 Enloe, C. 1983. *Does khaki become you? The militarization of women's lives*. Boston: South
31 End Press.
- 32 Fairhurst, G. T., & Snavely, B. K. 1983. Majority and token minority group relationships: Power
33 acquisition and communication. *Academy of Management Review*, 8(2): 292-300.
- 34 Farh, C. I. C., & Chen, G. 2018. Leadership and member voice in action teams: Test of a
35 dynamic phase model. *Journal of Applied Psychology*, 103(1): 97-110.
- 36 Fast, N. J., Burris, E. R., & Bartel, C. A. 2014. Managing to stay in the dark: Managerial self-
37 efficacy, ego defensiveness, and the aversion to employee voice. *Academy of Management*
38 *Journal*, 57(4): 1013-1034.
- 39 Feingold, A. 1994. Gender differences in personality: A meta-analysis. *Psychological*
40 *Bulletin*, 116(3): 429-456.
- 41 Finseraas, H., Johnsen, A. A., Kotsadam, A., & Torsvik, G. 2016. Exposure to female colleagues
42 breaks the glass ceiling-evidence from a combined vignette and field experiment. *European*
43 *Economic Review*, 90: 363-374.
- 44 Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. 2002. A model of (often mixed) stereotype
45 content: Competence and warmth respectively follow from perceived status and
46 competition. *Journal of Personality and Social Psychology*, 82(6): 878-902.
- 47 Fitts, P. M. 1954. The information capacity of the human motor system in controlling the
48 amplitude of movement. *Journal of Experimental Psychology*, 47: 381-391.
- 49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 Forgas, J. P., & Williams, K. D. (Eds.). 2001. *Social influence: Direct and indirect processes*.
4 Philadelphia; Psychology Press.
- 5 Gorenflo, D. W., & Crano, W. D. 1989. Judgmental subjectivity/objectivity and locus of choice
6 in social comparison. *Journal of Personality and Social Psychology*, 57(4): 605-614.
- 7 Halpern, D. F. 1986. A different answer to the question, "Do sex related differences in spatial
8 abilities exist?" *American Psychologist*, 41: 1014-1015.
- 9 Harrison, D.A., & Klein, K.J. 2007. What's the difference? Diversity constructs as separation,
10 variety, or disparity in organizations. *Academy of Management Review*, 32: 1199-1228.
- 11 Heilman, M. E. 2001. Description and prescription: How gender stereotypes prevent women's
12 ascent up the organizational ladder. *The Journal of Social Issues*, 57(4): 657-674.
- 13 Heilman, M. E. 2012. Gender stereotypes and workplace bias. *Research in Organizational*
14 *Behavior*, 32: 113-135.
- 15 Hirschfeld, R. R., Jordan, M. H., Feild, H. S., Giles, W. F., & Armenakis, A. A. 2005. Teams'
16 female representation and perceived potency as inputs to team outcomes in a predominantly
17 male field setting. *Personnel Psychology*, 58(4): 893-924.
- 18 Hofmann, D. A., & Gavin, M. B. 1998. Centering decisions in hierarchical linear models:
19 Implications for research in organizations. *Journal of Management*, 24: 623-641.
- 20 Holland, S. L. 2006. The dangers of playing dress-up: Popular representations of Jessica Lynch
21 and the controversy regarding women in combat. *Quarterly Journal of Speech*, 92: 27-50.
- 22 Hollenbeck, J. R., & Wright, P. M. 2017. Harking, sharking, and tharking: Making the case for
23 post hoc analysis of scientific data. *Journal of Management*, 43: 5-18.
- 24 Hollenbeck, J. R., Beersma, B., & Schouten, M. E. 2012. Beyond team types and taxonomies: A
25 dimensional scaling conceptualization for team description. *Academy of Management*
26 *Review*, 37: 82-106.
- 27 Hollenbeck, J. R., Ilgen, D. R., Segoe, D. J., Hedlund, J., Major, D. A., & Phillips, J. 1995.
28 Multilevel theory of team decision making: Decision performance in teams incorporating
29 distributed expertise. *Journal of Applied Psychology*, 80: 292-316.
- 30 Howell, T. M., Harrison, D. A., Burris, E. R., & Detert, J. R. 2015. Who gets credit for input?
31 Demographic and structural status cues in voice recognition. *Journal of Applied*
32 *Psychology*, 100: 1765-1784.
- 33 Janis, I. 1972. *Victims of groupthink: A psychological study of foreign policy decisions and*
34 *fiascoes*. Boston, MA: Houghton Mifflin.
- 35 Joshi, A., & Roh, H. 2009. The role of context in work team diversity research: A meta-analytic
36 review. *Academy of Management Journal*, 52(3): 599-627.
- 37 Kanter, R. A. 1977. *Men and women of the corporation*. New York: Harper.
- 38 Kelley, H. H. 1971. Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley,
39 R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*
40 *(pp. 1-26)*. New York: General Learning Press.
- 41 Kozlowski, S. W. J., & Ilgen, D. R. 2006. Enhancing the effectiveness of work groups and
42 teams. *Psychological Science in the Public Interest*, 7(3): 77-124.
- 43 LePine, J. A., Hollenbeck, J. R., Ilgen, D. R., Colquitt, J. A., & Ellis, A. 2002. Gender
44 composition, situational strength, and team decision-making accuracy: A criterion
45 decomposition approach. *Organizational Behavior and Human Decision Processes*, 88(1):
46 445-475.
- 47 LePine, J. A., & Van Dyne, L. 1998. Predicting voice behavior in work groups. *Journal of*
48 *Applied Psychology*, 83: 853-868.
- 49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 Li, A. N., Liao, H., Tangirala, S., & Firth, B. M. 2017. The content of the message matters: The
4 differential effects of promotive and prohibitive team voice on team productivity and safety
5 performance gains. *Journal of Applied Psychology*, 102(8): 1259-1270.
- 6 Liang, J., Farh, C. I. C., & Farh, J. L. 2012. Psychological antecedents of promotive and
7 prohibitive voice: A two-wave examination. *Academy of Management Journal*, 55: 71-92.
- 8 Liang, J., Shu, R., & Farh, C. I. C. 2018. Differential implications of team member promotive
9 and prohibitive voice on innovation performance in research and development project teams:
10 A dialectic perspective. *Journal of Organizational Behavior*.
- 11 Lin, N. 1999. Social networks and status attainment. *Annual Review of Sociology*, 25(1): 467-
12 487.
- 13 Lord, R. G. 1977. Functional leadership behavior: Measurement and relation to social power and
14 leadership perceptions. *Administrative Science Quarterly*, 22(1): 114-133.
- 15 Loyd, D. L., Wang, C. S., Phillips, K. W., & Lount, Robert B., Jr. 2013. Social category diversity
16 promotes premeeting elaboration: The role of relationship focus. *Organization
17 Science*, 24(3): 757-772.
- 18 Maass, A., & Clark, R. D. 1984. Hidden impact of minorities: Fifteen years of minority influence
19 research. *Psychological Bulletin*, 95(3): 428-450.
- 20 Maass, A., West, S. G., & Cialdini, R. B. 1987. Minority influence and conversion. *Review of
21 Personality and Social Psychology*, 8: 55-79.
- 22 Maier, N. R., & Solem, A. R. 1952. The contribution of a discussion leader to the quality of
23 group thinking: The effective use of minority opinions. *Human Relations*, 5: 277-288.
- 24 Maltz, D., & Borker, R. 1983. A cultural approach to male-female miscommunication. In J.
25 Gumperz (Ed.), *Language and Social Identity (Studies in Interactional Sociolinguistics, pp.
26 196-216)*. Cambridge: Cambridge University Press.
- 27 Mannix, E., & Neale, M. A. 2005. What differences make a difference? The promise and reality
28 of diverse teams in organizations. *Psychological Science in the Public Interest*, 6(2): 31-55.
- 29 Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. 2001. A temporally based framework and
30 taxonomy of team processes. *Academy of Management Review*, 26(3): 356-376.
- 31 Martin, R., Martin, P. Y., Smith, J. R., & Hewstone, M. 2007. Majority versus minority influence
32 and prediction of behavioral intentions and behavior. *Journal of Experimental Social
33 Psychology*, 43: 763-771.
- 34 Mayer, R. C., Davis, J. H., & Schoorman, F. D. 1995. An integrative model of organizational
35 trust. *Academy of Management Review*, 20: 709-734.
- 36 Maynes, T. D., & Podsakoff, P. M. 2014. Speaking more broadly: An examination of the nature,
37 antecedents, and consequences of an expanded set of employee voice behaviors. *Journal of
38 Applied Psychology*, 99: 87-112.
- 39 McClean, E., Martin, S. R., Emich, K. J., & Woodruff, T. 2018. The social consequences of
40 voice: An examination of voice type and gender on status and subsequent leader emergence.
41 *Academy of Management Journal*, 61: 1869-1891.
- 42 Mesmer-Magnus, J., & DeChurch, L. A. 2009. Information sharing and team performance: A
43 meta-analysis. *Journal of Applied Psychology*, 94(2): 535-546.
- 44 Michaels, J. 2013, February 3. Debate over women in combat shifts to physical strength. *USA
45 Today*. Retrieved from [http://www.usatoday.com/story/news/nation/2013/02/02/marines-
46 army-standards-women-combat/1884141/](http://www.usatoday.com/story/news/nation/2013/02/02/marines-army-standards-women-combat/1884141/)
- 47 Morgeson, F. P., DeRue, D. S., Karam, E. P. 2010. Leadership in teams: A functional approach
48 to understanding leadership structures and processes. *Journal of Management*, 36: 5-39.
- 49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 Morgeson, F. P., & Humphrey, S. E. 2006. The Work Design Questionnaire (WDQ): Developing
4 and validating a comprehensive measure for assessing job design and the nature of
5 work. *Journal of Applied Psychology*, 91: 1321-1339.
- 6 Morrison, E. W. 2011. Employee voice behavior: Integration and directions for future research.
7 *Academy of Management Annals*, 5: 373-412.
- 8 Morrison, E. W., & Milliken, F. J. 2000. Organizational silence: A barrier to change and
9 development in a pluralistic world. *Academy of Management Review*, 25(4): 706-725.
- 10 Morrison, E. W., Wheeler-Smith, S., & Kamdar, D. 2011. Speaking up in groups: A cross-level
11 study of group voice climate and voice. *Journal of Applied Psychology*, 96(1): 183-191.
- 12 Morrison, E. W. 2014. Employee voice and silence, *Annual Review of Organizational*
13 *Psychology and Organizational Behavior*, 1: 173-197.
- 14
15
16 Moscovici, S. 1980. Toward a theory of conversion behavior. *Advances in Experimental Social*
17 *Psychology*, 13: 209-239.
- 18 Moscovici, S. 1985. Social influence and conformity. In G. Lindzey, & E. Aronson (Eds.),
19 *Handbook of social psychology (Vol. 2, pp. 347-412)*. New York: McGraw-Hill.
- 20 Moscovici, S., Lage, E., & Naffrechoux, M. 1969. Influence of a consistent minority on the
21 responses of a majority in a color perception task. *Sociometry*, 32(4): 365-380.
- 22 Moscovici, S., Mucchi-Faina, A., & Maass, A. (Eds.). 1994. *Minority influence*. Chicago:
23 Nelson-Hall.
- 24 Moscovici, S., & Mugny, G. 1983. Minority influence. In P.B. Paulus (ED.), *Basic Group*
25 *Processes*, (pp. 41-65). New York: Springer Verlag.
- 26 Moskowitz, G. B. 1996. The mediational effects of attributions and information processing in
27 minority social influence. *The British Journal of Social Psychology*, 35: 47-66.
- 28 Mugny, G. 1982. *The power of minorities*. London: Academic Press.
- 29 Mugny, G., Butera, F., Sanchez-Mazas, M., & Pe´rez, J. A. 1995. Judgements in conflict: The
30 conflict elaboration theory of social influence. In B. Booth, R. Hirsig, A. Helminger, B.
31 Meier, & R. Volkart (Eds.), *Perception evaluation interpretation (pp. 160-168)*. Gottingen:
32 Hogrefe & Huber.
- 33 Nebus, J. 2006. Building collegial information networks: A theory of advice network
34 generation. *Academy of Management Review*, 31: 615-637.
- 35 Nemeth, C. J. 1986. Differential contributions of majority and minority influence. *Psychological*
36 *Review*, 93(1): 23-32.
- 37 Nemeth, C. J., & Kwan, J. L. 1985. Originality of word associations as a function of majority vs.
38 minority influence. *Social Psychology Quarterly*, 48(3): 277-282.
- 39 Nemeth, C. J., & Kwan, J. L. 1987. Minority influence, divergent thinking and detection of
40 correct solutions. *Journal of Applied Social Psychology*, 17(9): 788-799.
- 41 Nemeth, C. J., & Wachtler, J. 1983. Creative problem solving as a result of majority vs minority
42 influence. *European Journal of Social Psychology*, 13(1), 45-55.
- 43 Nemeth, C. J., Connell, J. B., Rogers, J. D., & Brown, K. S. 2001. Improving decision making by
44 means of dissent. *Journal of Applied Social Psychology*, 31(1): 48-58.
- 45 Nemeth, C. J., & Nemeth-Brown, B. 2003. Better than individuals? The potential benefits of
46 dissent and diversity for group creativity. In P. B. Paulus, & B. A. Nijstad (Eds.), *Group*
47 *creativity: Innovation through collaboration; group creativity: Innovation through*
48 *collaboration (pp. 63-84)*. Oxford University Press, New York, NY.
- 49 Nemeth, C., & Rogers, J. 1996. Dissent and the search for information. *British Journal of Social*
50 *Psychology*, 35(1): 67-76.
- 51
52
53
54
55
56
57
58
59
60

- 1
2
3 Nemeth, C., Mosier, K., & Chiles, C. 1992. When convergent thought improves performance:
4 Majority versus minority influence. *Personality and Social Psychology Bulletin*, 18(2): 139-
5 144.
- 6 Perlow, L., & Williams, S. 2003. Is silence killing your company? *Harvard Business Review*,
7 81: 52-58.
- 8 Phillips, K. W. 2003. The effects of categorically based expectations on minority influence: The
9 importance of congruence. *Personality and Social Psychology Bulletin*, 29(1): 3-13.
- 10 Phillips, K. W., & Loyd, D. L. 2006. When surface and deep-level diversity collide: The effects
11 on dissenting group members. *Organizational Behavior and Human Decision*
12 *Processes*, 99(2): 143-160.
- 13 Phillips, K. W., Liljenquist, K. A., & Neale, M. A. 2009. Is the pain worth the gain? The
14 advantages and liabilities of agreeing with socially distinct newcomers. *Personality and*
15 *Social Psychology Bulletin*, 35(3): 336-350.
- 16 Phillips, K. W., Mannix, E. A., Neale, M. A., & Gruenfeld, D. H. 2004. Diverse groups and
17 information sharing: The effects of congruent ties. *Journal of Experimental Social*
18 *Psychology*, 40(4): 497-510.
- 19 Phillips, K. W., Northcraft, G. B., & Neale, M. A. 2006. Surface-level diversity and decision-
20 making in groups: When does deep-level similarity help? *Group Processes & Intergroup*
21 *Relations*, 9(4): 467-482.
- 22 Plaut, V. C. (2010). Diversity science: Why and how difference makes a difference.
23 *Psychological Inquiry*, 21: 77-99.
- 24 Premeaux, S. F., & Bedeian, A. G. 2003. Breaking the silence: The moderating effects of self-
25 monitoring in predicting speaking up in the workplace. *Journal of Management*
26 *Studies*, 40(6): 1537-1562.
- 27 Raudenbush, S. W., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. 2004. *HLM 6:*
28 *Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software
29 International.
- 30 Ridgeway, C. L., & Correll, S. J. 2004. Unpacking the gender system: A theoretical perspective
31 on gender beliefs and social relations. *Gender & Society*, 18(4): 510-531.
- 32 Rosette, A. S., & Tost, L. P. 2010. Agentic women and communal leadership: How role
33 prescriptions confer advantage to top women leaders. *Journal of Applied Psychology*, 95(2):
34 221-235.
- 35 Rudman, L. A., & Glick, P. 2001. Prescriptive gender stereotypes and backlash toward agentic
36 women. *Journal of Social Issues*, 57(4): 743-762.
- 37 Rustad, Michael. 1982. *Women in Khaki: The American Enlisted Woman*. New York: Praeger.
- 38 Sandberg, S. 2013. *Lean in: Women, work, and the will to lead*. Random House.
- 39 Sandberg, S. & Grant, A. 2015, January 11. Speaking while female. *The New York Times*.
40 Retrieved from [https://www.nytimes.com/2015/01/11/opinion/sunday/speaking-while-](https://www.nytimes.com/2015/01/11/opinion/sunday/speaking-while-female.html)
41 [female.html](https://www.nytimes.com/2015/01/11/opinion/sunday/speaking-while-female.html)
- 42 Schwenk, C. R. 1990. Conflict in organizational decision making: An exploratory study of its
43 effects in for-profit and not-for-profit organizations. *Management Science*, 36: 436-448.
- 44 Shrout, P. E., & Fleiss, J. L. 1979. Intraclass correlations: Uses in assessing rater
45 reliability. *Psychological Bulletin*, 86(2): 420-428.
- 46 Statistic Brain. 2017. Women in the military statistics. Retrieved from
47 <https://www.statisticbrain.com/women-in-the-military-statistics/>
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 Tessier, M. 2016, October 27. Speaking while female, and at disadvantage. *The New York*
4 *Times*. Retrieved from [https://www.nytimes.com/2016/10/27/upshot/speaking-while-female-](https://www.nytimes.com/2016/10/27/upshot/speaking-while-female-and-at-a-disadvantage.html)
5 [and-at-a-disadvantage.html](https://www.nytimes.com/2016/10/27/upshot/speaking-while-female-and-at-a-disadvantage.html)
6
7 Thomas-Hunt, M. C., Ogden, T. Y., & Neale, M. A., 2003. Who's really sharing? Effects of
8 social and expert status on knowledge exchange within groups. *Management Science*, 49(4)
9 464-477.
- 10 U. S. Department of Labor. 1977. *Dictionary of occupational titles (4th ed.)*. Washington, DC:
11 U.S. Government Printing Office.
- 12 Volpato, C., Maass, A., Mucchi-Faina, A., & Vitti, E. 1990. Minority influence and social
13 categorization. *European Journal of Social Psychology*, 20(2), 119-132.
- 14 van Dijk, H., Meyer, B., van Engen, M., & Loyd, D. 2016. Microdynamics in diverse teams: A
15 review and integration of the diversity and stereotyping literatures. *Academy of Management*
16 *Annals*, 11: 517-557.
- 17
18 van Knippenberg, D., De Dreu, C. K., & Homan, A. C. 2004. Work group diversity and group
19 performance: An integrative model and research agenda. *Journal of Applied Psychology*, 89:
20 1008-1022.
- 21
22 van Knippenberg, D., & Mell, J. N. 2016. Past, present, and potential future of team diversity
23 research: From compositional diversity to emergent diversity. *Organizational Behavior and*
24 *Human Decision Processes*, 136: 135-145.
- 25 Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of*
26 *Psychology*, 58: 515-541.
- 27 Westphal, J. D., & Milton, L. P. (2000). How experience and network ties affect the influence of
28 demographic minorities on corporate boards. *Administrative Science Quarterly*, 45(2): 366-
29 398.
- 30
31 Woodworth, R. S. 1899. Accuracy of voluntary movement. *The Psychological Review:*
32 *Monograph Supplements*, 3(3): 1-114.
- 33 Yoder, J. D., Adams, J., & Prince, H. T. 1983. The price of a token. *Journal of Political and*
34 *Military Sociology*, 11(2): 325-337.
- 35 Zaccaro, S. J., Rittman, A. L., & Marks, M. A. 2001. Team leadership. *Leadership Quarterly*,
36 12(4): 451-483.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

[†] 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).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

TABLE 2 REGRESSION RESULTS PREDICTING TOKEN FEMALE (YELLOW HELMET) VOICE ENACTMENT

	Yellow helmet voice enactment			
	<u>M1</u>	<u>M2</u>	<u>M3</u>	<u>M4</u>
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 $-2 \times$ 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 $-2 \times$ 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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURE 1

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

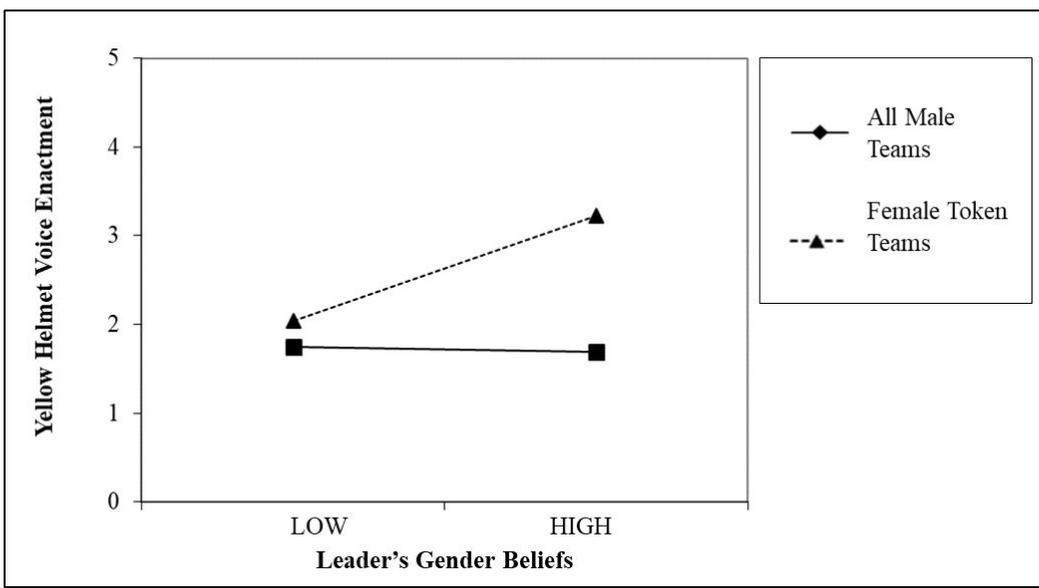
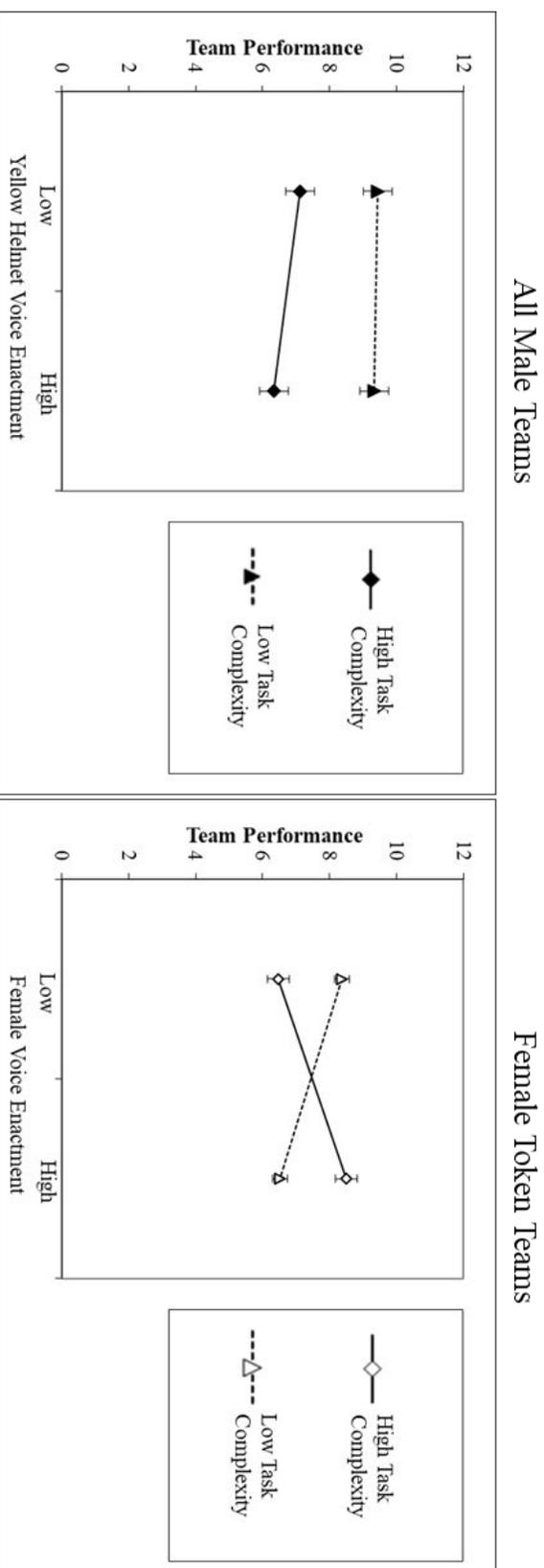


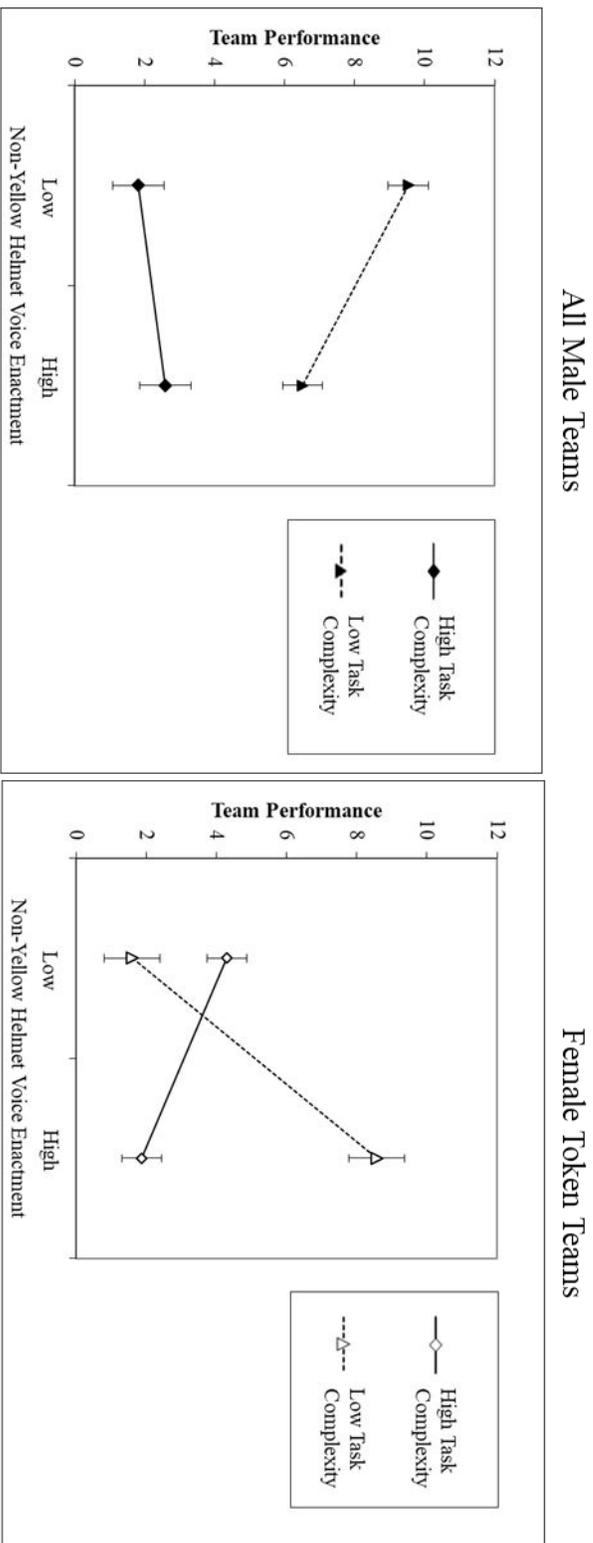
FIGURE 2

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



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

AUTHOR BIOGRAPHIES

Crystal I. C. Farh (farh@uw.edu) is an associate professor of management and organization at the University of Washington. She received her Ph.D. from the University of Maryland and previously served on the faculty of Michigan State University. Her research interests focus on employee voice, teams and leadership, and managing cross-cultural interfaces.

Jo K. Oh (kyoungjo.oh@uconn.edu) is an assistant professor at the University of Connecticut, School of Business. He received his Ph.D. from Michigan State University. His current research focuses on abusive supervision, teams, emotions, and motivation.

John R. Hollenbeck (jrh@broad.msu.edu) holds the positions of University Distinguished Professor at Michigan State University and Eli Broad Professor of Management at the Eli Broad Graduate School of Business Administration. Dr. Hollenbeck received his PhD in Management from New York University in 1984.

Andrew Yu (andrew.yu@unimelb.edu.au) is an assistant professor with the Faculty of Business and Economics at the University of Melbourne, Australia. He received his Ph.D. from Michigan State University and his M.A. in Economics and B.A. in Business Administration from California State University, Fullerton. His research focuses on interpersonal relationships, group dynamics, and leadership.

Stephanie M. Lee (stephanie_kunst@baylor.edu) is a clinical assistant professor at the Hankamer School of Business at Baylor University. Her current research interests include team dynamics, leadership, and voice.

Danielle D. King (danielle.d.king@rice.edu) is an assistant professor of Psychological Sciences at Rice University. She received her M.A. and Ph.D. in Organizational Psychology from Michigan State University and her B.A. from Spelman College. Her research focuses on employee resilience, identity intersectionality and authenticity, and workplace citizenship behavior.