GENDER DIVERSITY, CONFLICT, AND PSYCHOLOGICAL SAFETY IN TEAMS: A CONCEPTUAL AND EMPIRICAL EXAMINATION

By

Emily Redler

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Arts in
Psychology-Industrial & Organizational

Charlotte

2022

Approved by:

_____
Dr. David Woehr

_____
Dr. Linda Shanock

_____
Dr. Jill Yavorsky

ABSTRACT

EMILY REDLER. Gender Diversity, Conflict, and Psychological Safety: A Conceptual and Empirical Examination.
(Under the direction of DR. DAVID WOEHR)

Teamwork is common in daily work life, and it is likely that, when working within teams, conflict may occur. Conflict negatively affects a team's psychological safety – the feeling of safety and willingness to share one's ideas and take risks while in a team. The gender diversity of a team may moderate this negative relationship. In the present study, I examined the effects of team-level task and relationship conflict on team-level psychological safety, moderated by team gender diversity. Results showed a negative main effect of both types of conflict on psychological safety. However, gender diversity was not found to be a significant moderator. I then examined the proportion of women on a team as an exploratory moderation analysis and found that a team's proportion of women strengthened the negative relationship between task conflict and psychological safety – the opposite direction of moderation predicted. A discussion follows of a potential backlash effect of women on a team toward other women, exacerbating the already negative effects of task conflict. Future directions and limitations are discussed.

KEYWORDS: Psychological safety, conflict, gender diversity, teamwork, backlash

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor, Dr. David Woehr, and thank him for his assistance and expertise, both of which greatly aided me in executing this project. Additionally, I would like to express my gratitude to my committee members, Dr. Linda Shanock and Dr. Jill Yavorsky, both of whom went above and beyond their responsibilities as committee members to help me in developing and refining this thesis. Lastly, I would like to acknowledge Dr. Matthew Ohland, Dr. Misty Loughry, and again Dr. David Woehr, founders of the excellent CATME program, for allowing me use of their data; without this CATME data, my project would not have been achievable, so I am appreciative of their generosity.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Teamwork is a demand in many facets of daily life, from sports to the workplace; as such, organizational scientists often study teamwork and team processes like conflict and psychological safety. These processes do not always operate independent of context, though, so it is important to consider factors (such as team composition) that may influence these processes' effects. For instance, the gender breakdown of a team may affect how the team is able to constructively deal with conflict. However, gender diversity as a contextual moderator has not often been examined in the teams literature. For the present study, I examined the relationship between team conflict and team psychological safety moderated by team gender diversity.

LITERATURE REVIEW

Psychological safety is a "shared belief held by members of a team that the team is safe for interpersonal risk taking" (Edmondson, 1999, p. 350). Although often studied at the individual level in organizations (Frazier et al., 2017), it is highly relevant to teams (e.g., Edmondson, 1999; Kessel et al., 2012). Feeling safe to take risks and voice one's ideas is imperative in a team setting; a lack of psychological safety has been associated with negative outcomes such as lowered team performance (Kostopoulos & Bozionelos, 2011). In addition, psychological safety has been shown to be a mediator of the relationship between team processes (e.g., context support and coaching) and team performance (Edmondson, 1999).

In a meta-analysis by Frazier et al. (2017), individual psychological safety was related to a variety of both antecedents and outcomes, including emotional stability, autonomy, interdependence, organizational citizenship behaviors, and commitment. We know less about psychological safety at the group (team) level, however; though Frazier et al. (2017) found individual- and group-level psychological safety to be related to a similarly large number of antecedents (e.g., learning orientation, leader trust, autonomy, interdependence, peer support, and organizational support) and outcomes (e.g., engagement, task performance, information sharing, and satisfaction), there were significantly fewer studies (*k*) examining group-level constructs than individual-level constructs. In addition, when examining group-level constructs that *directly* deal with team dynamics, only interdependence and peer support were found to be significantly related to psychological safety, and both had positive effects.

Conflict in teams, on the other hand, may have detrimental effects; thus, conflict could get in the way of teams feeling psychologically safe. Given that conflict in teams is a widely

studied concept with numerous implications, a logical next step for the psychological safety literature would be to examine the role of team conflict in relation to team psychological safety.

In addition, with the recent increasing push for diversity in teams, it makes sense to examine how increasing diversity may affect team dynamics. Specifically, it is important to explore whether the level of diversity in teams enhances or detracts from team conflict's potentially negative effects on team psychological safety. While diversity of all types has been under the spotlight in recent years, for the current paper, gender diversity is the focus given the plethora of literature on gender differences in conflict and conflict management styles (e.g., Brewer et al., 2002; Rahim & Katz, 2019).

In sum, my goals for this study were to add to the teams literature by examining the relationships between various forms of conflict and team psychological safety and to explore how gender diversity within a team may affect these relationships.

Team Conflict

The antecedents for this study were various forms of team conflict. Colloquially, the term *conflict* is unspecific. In psychological literature, however, conflict can be more specifically defined with terms such as intergroup conflict (conflict amongst groups), inter-role conflict (in which an individual has multiple roles that conflict), and inter-organization conflict (conflict amongst organizations). Each of these more specific forms of conflict may fit the general definition given by Wall & Callister (1995): conflict is "a process in which one party perceives that its interests are being opposed or negatively affected by another party" (p. 517).

Given the pervasiveness and necessity of teamwork throughout one's life (e.g., Salas et al., 2000), conflict within teams – team conflict – is unavoidable. Team conflict may appear when an individual feels negatively affected by another individual or group of individuals within

their team; differences in ideology, power, or resource availability are a few of the many reasons this can occur (De Dreu & Van De Vliert, 1997). In addition, team members may conflict over issues unique to teamwork itself, such as social loafing or how to distribute work effectively (Wageman, 1995). Team conflict, however, is not a single construct. In her seminal work on team conflict, Jehn (1995) conceptualized two types of team conflict: relationship conflict and task conflict.

Jehn (1995) described relationship conflict as existing when there are "interpersonal incompatibilities among group members, which typically includes tension, animosity, and annoyance among members within a group" (p. 258). When group members fail to create positive relationships and instead reside in a team rife with personal conflicts, there may be negative effects on team-level outcomes. For example, relationship conflict has been shown to be negatively associated with team performance, team satisfaction, and team effectiveness (e.g., De Dreu & Van Vianen, 2001; De Dreu & Weingart, 2003; Jehn, 1995).

Relationship conflict may impact these team-level outcomes through trust, as higher levels of relationship conflict within a team may negatively affect trust (Lau & Cobb, 2010). Trust is imperative for team decision-making, as a lack of trust amongst team members encourages these individuals to rely on their initial negative perceptions of other team members rather than on built relationships (Lau & Cobb, 2010). These negative attributions then could lead to poor decision-making (Bono et al., 2002; de Wit et al., 2013). Further, high levels of relationship conflict have been shown to negatively affect team members' acceptance of group decisions (Simons & Peterson, 2000).

Task conflict, on the other hand, is defined by Jehn (1995) as existing "when there are disagreements among group members about the content of the tasks being performed, including

differences in viewpoints, ideas, and opinions" (p. 258). Jehn found that task conflict was negatively related to satisfaction and intent to remain in the group, though it did seem to help with decision-making and performance. Further research complicates these findings; De Dreu and Weingart (2003), for instance, found that the average correlation between task conflict and group performance was -.23.

Perspectives on Conflict

The literature surrounding team conflict and its various forms indicates a series of complex and nuanced relationships. For instance, though team conflict may seem to be an entirely negative phenomenon that organizations should strive to avoid, it has also been hypothesized to have some positive effects. Multiple scholars have even gone so far as to say some level of conflict in teams is necessary for optimal outcomes (e.g., Caudron, 1998; De Dreu & Van De Vliert, 1997).

Further, task and relationship conflict themselves, though distinct constructs, may be intertwined. For instance, Gamero et al. (2008) examined the relationships between team relationship conflict, task conflict, and affect in teams of bank employees. They found that relationship conflict fully mediated the negative relationship between team task conflict and team affect. In addition, results showed that the level of interaction among team members moderated the relationship between task conflict and relationship conflict; lower levels of interaction amongst team members strengthened the positive relationship between task and relationship conflict, and vice versa.

This evidence contradicts the separation perspective discussed by Janssen et al. (1999). The separation perspective on task and person (i.e., relationship) conflict supposes that these two types of conflict are independent and exert independent effects on team outcomes. Even early

conflict scholars, however, contradict this perspective, citing how these types of conflict overlap in many ways (e.g., Gamero et al., 2008).

As an alternative to the separation perspective, Janssen et al. (1999) proposed a complexity perspective in which task and person conflict co-exist in teams and are often mistaken for each other but should be managed differently. For instance, Costa et al. (2015) examined research teams' conflict, performance, and engagement. Results showed a positive main effect of team resources on work engagement, but this effect was moderated by conflict; the effect was strengthened by higher levels of task conflict and weakened by higher levels of relationship conflict. These findings support the notion that task conflict may have a positive effect on certain team outcomes, with relationship conflict having the opposite effect.

The complexity perspective informs a great deal of research on conflict management and recommendations for organizations, creating a "two sides of the coin" recommendation: organizations should encourage task conflict and mitigate relationship conflict during team decision making (Janssen et al., 1999). However, this simple perspective of relationship conflict being negative and task conflict being positive oversimplifies their effects, as indicated by the conflicting findings on this topic (e.g., Auh et al., 2014; De Dreu & Weingart, 2003; Jehn, 1995). Varying perspectives – such as that of Gamero et al. (2008) above – have emerged in recent years, attempting to make sense of the often conflicting and contradictory research on team conflict. Other studies have echoed their findings that task and relationship conflict are not completely independent, though they are distinct (e.g., Friedman et al., 2000; Medina et al., 2005). For the present study, I employ this more nuanced perspective of relationship and task conflict as distinct yet related constructs.

Psychological Safety and Conflict

After defining my antecedents (relationship and task conflict) and outcome (psychological safety), it's important to examine literature on the relationship between the two. Very few studies have examined the direct effect of conflict on team psychological safety. Instead, many studies examine these forms of conflict as moderators; for instance, Kostopoulos and Bozionelos (2011) examined project teams' learning and task conflict. They found that task conflict moderated the relationship between psychological safety and exploitative learning such that, for teams with higher levels of task conflict, this relationship was strengthened, and vice versa.

Wilkens and London (2006) conducted one of the few studies examining the direct effect of conflict on psychological safety. In their study, groups of participants in a hospital self-disclosed their ratings of psychological safety and task conflict within their team. Results showed that teams with more task conflict tended to have lower levels of psychological safety. This informs my hypotheses relating to conflict and psychological safety; a negative interpersonal climate may be associated with lower psychological safety.

The Influence of Gender Diversity

An additional contribution of my study is an examination of gender diversity's effect on the relationships described above. Like teamwork, gender is an inescapable construct in today's society, affecting nearly all facets of life (e.g., Acker, 1990). Due to the ubiquity of these topics, a variety of literature has examined them in conjunction, most commonly in direct relation; for instance, the main effects of a team's gender diversity on various team processes are commonly examined in team and diversity literatures (e.g., Fenwick & Neal, 2001; Myaskovsky et al., 2005). However, findings in this realm have been mixed, potentially indicating a more nuanced (and potentially indirect) relationship than has been previously examined.

Fenwick and Neal (2001), for example, investigated the effect of group gender composition on performance. The researchers examined how groups with gender homogeneity and heterogeneity performed on a business simulation and found that groups with gender parity or a female majority performed better than groups with a male majority. Fenwick and Neal theorized that this difference in performance was mainly due to women's inherently more interactive, people-oriented, and cooperative work styles being more effective than men's inherently analytical and competitive work styles. Similarly, McLeod and colleagues (1996) found that team diversity is associated with not only general team performance but also quality of work. In their study, they found that more ethnically diverse groups were likely to create ideas that were judged to be more effective and feasible, contributing to a judgment of overall higher quality. Although this study varied ethnic diversity rather than gender diversity, it may be argued that similar effects would hold for heterogenous and homogenous gender groups; this assertion is supported by additional literature on the topic (e.g., Bear and Woolley, 2011) as well as by Status Characteristics Theory (SCT) (Berger et al., 1972; Ridgeway, 2014).

Based on SCT, group members develop expectations for both high- and low-status members of their group based on beliefs formed through prior experience, contributing to the effects we see in diversity literature (e.g., McLeod et al., 1996). This logic holds for a variety of different diffuse status characteristics – characteristics that are differentially evaluated and have both specific and general expectations for performance (Savage et al., 2020); therefore, by the logic of SCT, McLeod et al.'s findings varying ethnic diversity – a diffuse status characteristic – may also hold when varying gender diversity, another example of a diffuse status characteristic.

Alternately, some studies have found negative main effects of gender diversity. For example, evidence shows that women in a mixed-gender group may be perceived to be less

competent and less influential than men in these groups (Heilman & Welle, 2006). Women feeling less powerful and less able to contribute to a group, and their contributions themselves being perceived as less competent, may deprive a group of potential benefits gender diversity can bring (e.g., Fenwick & Neal, 2001, McLeod et al., 1996). Even if women are able to contribute their diverse opinions to a team, these varied perspectives may bring other negative effects. For instance, van den Oever and Beerens (2021) found that teams with more gender diversity tended to have more varied information and perspectives, leading to more task conflict.

The mixed and even contradictory findings relating to the effects of gender diversity on team processes indicate potential utility in examining gender from an alternative perspective. Specifically, gender diversity may function as a key contextual variable with respect to team processes. That is, gender diversity may serve to moderate the relationship among key team process inputs and outcomes; in this case, gender diversity of a team may moderate the relationship between team conflict and team psychological safety.

Gender, Conflict, and Status Characteristics Theory

The notion that gender may influence this relationship stems from the fact that men and women socialize differently in society and are met with differing expectations of their behavior. These differing expectations are explained via SCT (Berger et al., 1972), discussed above. In short, via SCT groups develop expectations for others based their diffuse status characteristics. In order for a diffuse status characteristic to affect group dynamics, though, it must be activated through salience; in essence, the diffuse status characteristic must differentiate between people and be relevant in a certain situation.

In addition, according to SCT, one must prove that a status characteristic is *not* relevant in a given situation, rather than prove that it is, in order for others to not assign them expectations

based upon that characteristic; unless otherwise proven (usually via prior experience), a status characteristic will be assumed to be relevant, and the expectations associated with it will be applied. Assuming a diffuse status characteristic has been activated and has not been proven to be irrelevant in a given situation, an individual will assign performance expectations ("states", as Berger and colleagues say) to themselves and to other individuals consistently. In practice, this may manifest in a large variety of situations; for instance, Myaskovsky and colleagues (2005) conducted a study of small group workstyles, varying group gender compositions. They found that found that women in mixed-gender groups – specifically mixed-gender groups in which they were in the gender minority – were less likely to talk than men in these groups, suggesting a lack of power women may hold in these gendered interactions.

How, though, this power imbalance came to be may be explained with SCT. Applying the assumptions of SCT, participants in these mixed-gender groups did not have explicit information that would lead them to be sure that the diffuse status characteristic being examined – in this case, gender – was irrelevant to the situation. With this assumption being fulfilled, group members were able to assign performance expectations to themselves and the others in their group. Stereotypically, the performance expectation to be assigned to women would be lower than that assigned to men, and there may even be specific expectations of women speaking less than men. Therefore, the men in the group would assign themselves higher performance expectations and women lower, and because these men made up the majority of the group, these expectations would dominate and permeate to the rest of the group.

Going further, these performance expectations an individual assigned to oneself and to others – in a situation where the diffuse status characteristic in question is salient and has not been proven to be irrelevant – will grant this individual power over the others in their group. This

further explains the findings of Myaskovsky and colleagues (2005) discussed above. The men in the mixed-gender groups in that study – who had the majority over the women – assigned lower performance expectations to the women in their group than to themselves. This directly gave the men power over the women in the group, potentially lowering women's participation. Not only were the women conforming to these constrictive performance expectations, but they may also have been directly dominated in conversations by men – people who, in this situation, had power over them.

Used in this manner, SCT may then explain a great deal of research that focuses on men's domination over women in mixed-gender teams. For example, there is a well-cited pattern of men dominating women in conversations. Anderson and Leaper (1998), for example, conducted a meta-analysis comparing men's and women's interruption patterns during conversations. They found men are more likely to interrupt women than vice versa and the types of interruptions men tend to use are intrusive (i.e., interrupting a speaker while they are speaking with the intent of taking over the conversation), suggesting a dominating motive. Following this, Brescoll (2011) conducted a series of studies examining gender and power in the context of speaking. In their first study, Brescoll (2011) examined U.S. senators' relative power and talking behavior, finding that senators with more power tended to talk more, and that these more powerful senators tended to be men. Men are able to hold and maintain power and dominance in situations via SCT and its assumptions, assigning constrained performance expectations to women.

In addition, this domination may fit within literature regarding conflict management styles; according to research, men tend to use more dominating conflict management styles and women tend to use more avoiding or compromising conflict management styles (e.g., Brewer et al., 2002; Rahim & Katz, 2019). In essence, where men enter conflict with dominance, women

enter with deference. Having a mix of dominating and deferring styles on a team may lessen the effects of conflict, as these complimentary conflict management styles will allow some to take the lead while others fall back. If a team consisted entirely of dominating conflict management styles, for instance, said conflict may be heightened.

There is additional evidence that having both men and women on a team may, at the very least, not exacerbate conflict. According to the gender similarities hypothesis (Hyde, 2005), men and women are more psychologically similar than they are different. Based on similarity perspectives, it could be hypothesized that higher levels of similarity among team members may lead to higher levels of interpersonal affect (Tajfel, 1981). Ultimately, this perspective suggests that the conflict/psychological safety relationship might be attenuated in gender homogenous teams.

To date, however, very few studies have examined the potential moderating effects of gender diversity, and only one of the relationship between conflict and psychological safety. Lee et al. (2018) empirically examined the moderating effect of gender diversity on the relationship between status conflict (conflict among team members over status positions; Bendersky & Hays, 2012) and psychological safety. Results based on simulated team scenarios indicated that, while status conflict was associated with lower psychological safety in general, this effect was lessened in teams with more gender diversity as opposed to gender dominant teams.

As a conceptual basis for their hypotheses, Lee et al. (2018) draw on evolutionary theory. Specifically, this evolutionary perspective draws on literature that has found that individuals on gender-diverse teams may exhibit more prosocial behaviors (e.g., altruism) intended to make them seem more attractive to other members. The premise of this argument is that people signal their good character and attract others via displays of altruism, as altruistic individuals are more

desirable as long-term partners than individuals who do not display altruistic qualities (Barclay, 2010).

Similarly, team members on a gender-diverse team may exhibit higher levels of cooperation. Farrelly (2011) conducted a study of male and female participants' attraction to one another as related to various examples of cooperation. Results showed that perceptions of cooperative behavior were important in mate selection. In addition, Ortmann and Tichy (1999) found that women are more cooperative in mixed-gender groups than in single-gender groups.

In addition to cooperation, individuals in gender-diverse teams may interact with more interpersonal sensitivity and thus manifest better problem-solving skills. Williams and Polman (2015) conducted a study of 202 management consultants and found that the presence of women in a situation increases one's likelihood to interact with interpersonal sensitivity, and these effects were found at both the individual-level (e.g., interacting one-on-one with a female employee) and team-level (i.e., in a mixed-gender team). In addition, having a high proportion of women on a team has been associated with better team problem solving (Hirschfeld et al., 2005), potentially indicating less conflict.

From the findings above, then, team members in a gender-diverse team may signal their attraction to potential mates via interacting with high levels of cooperation, altruism, and interpersonal sensitivity, all of which should theoretically be associated with a lessening of the negative effects of conflict on psychological safety, supporting Lee et al.'s (2018) findings and the hypotheses of the present study.

Hypotheses

In the present study, I examined the relationship between both task and relationship conflict and psychological safety in teams. In addition, I aimed to examine the potential moderating effect of gender diversity on this relationship.

Hypothesis 1: Team-level relationship conflict will be negatively related to team-level perceptions of psychological safety.

Hypothesis 2: Team-level task conflict will be negatively related to team-level perceptions of psychological safety.

Hypothesis 3a: Team gender diversity will moderate the relationship between team-level relationship conflict and team-level perceptions of psychological safety such that teams with higher levels of gender diversity will have a weaker relationship between team relationship conflict and team psychological safety.

Hypothesis 3b: Team gender diversity will moderate the relationship between team-level task conflict and team-level perceptions of psychological safety such that teams with higher levels of gender diversity will have a weaker relationship between team task conflict and team psychological safety.

METHOD

Participants and Procedure

For the present study, I used data collected via the Comprehensive Assessment of Team Member Effectiveness (CATME) program. The CATME program was developed in 2007 as a peer evaluation tool for examining team member effectiveness (Loughry et al., 2007) and currently includes measures of competence, student demographics, and various team processes such as task conflict, relationship conflict, and psychological safety. Since its inception, CATME has grown to be used by 1.6 million students at more than 2,500 institutions and is supported by nine articles in peer-reviewed journals and 88 conference presentations, papers, or workshops (info.catme.org).

I examined participants drawn from U.S. college students who used the CATME system for peer evaluation while engaging in teamwork-based projects between 2014 and 2021. Team members completed comprehensive evaluations of their team and teammates using the online CATME assessment system. Demographic measures were taken at the time of student team creation, and all team process measures were taken in one sitting at the end of each team's time working together.

It is important to note that, as briefly mentioned above, these data were collected by instructors to facilitate team formation and peer evaluation, i.e., the data were not collected solely for research purposes and I as the researcher did not form these teams. Therefore, teams included in the present study were actual project teams whose work had contributed to a grade in their course.

To ensure consistency in team size and composition, I implemented a number of filters on the larger CATME dataset. This filtering was done in R (R Core Team, 2021) and RStudio

(RStudio Team, 2021) using the "filter" function in the dplyr package (Wickham et al., 2021). Any teams with fewer than three or more than ten members were excluded from analysis, as well as any teams who included members identifying as anything other than a college student (i.e., they indicated they were either a freshman, sophomore, junior, or senior), any teams that worked together for less than 30 days (to ensure teams were adequately familiar with each other), and any teams in which any team member did not report their gender. Secondly, any duplicate entries (i.e., any entries between which the team ID, time point, and rater ID are all identical) or entries in which participants rated themselves (i.e., if the rater and ratee ID are identical) were removed from the dataset.

In addition, to ensure no bias relating to individuals who may not have completed all measures, all descriptive measures were calculated for the dataset both filtered and not filtered for only teams with entirely complete cases in terms of the variables of interest (see Appendix A Table 1A for descriptives without this filter). Results showed no significant difference from the dataset without these incomplete teams removed, so all further analyses were done using this filtered dataset for the sake of parsimony. After implementing these filters, the dataset for analysis had N = 1,030 participants in 255 teams.

Measures

Gender Diversity. Gender diversity within a team was assessed using a form of Blau's (1977) index, a widely used diversity metric (Lee et al., 2018). In this context, the index was used to determine if relationships between psychological safety and various forms of conflict differ by team gender breakdown. Blau's index calculates the diversity of a team based on the proportion of individuals in each category of the diversity characteristic. As gender for this analysis only has two categories (man and woman), in this context Blau scores would

theoretically only range from 0.5 (more gender-diverse) to 0 (more gender-homogenous). The equation for Blau's index is:

$$Blau = 1 - \sum_{i=1}^{k} p_i^2 \quad (1)$$

where p_i indicates the proportion of each category i in the group and k indicates the number of categories.

However, Blau's original (1977) index is limited when used to calculate diversity for variables with few categories, and has been shown to include bias based on the size and number of groups being used in the calculation (Biemann & Kearney, 2010). Equation 1 – that of Blau's original index – does not consider group size nor the number of categories being summed. To remedy this, Biemann and Kearney (2010) suggest using what they call $Blau_N$, a bias-corrected form of Blau taking into account group size and category number. The equation for this corrected form of Blau is:

$$Blau_N = 1 - \sum \frac{N_i(N_i - 1)}{N(N - 1)} \quad (2)$$

with N_i being the number of individuals in each category i, and N being the total number of individuals in the group. The results of this bias-corrected index, however, may still not be on a scale from 0 to 1. To standardize this index for ease of interpretation, I divided the result of Equation 2 by the maximum value obtained. This bias-corrected and standardized Blau's index ($Blau_{NS}$) will be used in all below analyses where a Blau's index is required. See Table 1 for correlations among the various iterations of Blau calculated here.

Relationship Conflict. Relationship conflict was assessed using the scale developed by Jehn & Manix (2001). The scale consists of three items: 1) "How much relation tension is there in your work group?", 2) "How often do people get angry while working in your group?", and 3) "How much emotional conflict is there in your work group?" (Jehn & Mannix, 2001). Students rated their perceived team relationship conflict via a 5-point Likert scale with 1 being "None or not at all" and 5 being "Very much or very often".

Task Conflict. Relationship conflict was assessed using the scale developed by Jehn & Manix (2001). The scale consists of three items: 1) "How much conflict of ideas is there in your work group?", 2) "How frequently do you have disagreements within your work group about the task of the project you are working on?", and 3) "How often do people in your work group have conflicting opinions about the project you are working on?" (Jehn & Mannix, 2001). Students rated their perceived team task conflict via a 5-point Likert scale with 1 being "None or not at all" and 5 being "Very much or very often".

Psychological Safety. Psychological safety was assessed using the scale developed by Edmondson (1999). The scale consists of seven items (three of which are reverse-worded), including "It is safe to take a risk on this team" and "Working with members of this team, my unique skills are valued and utilized" (Edmondson, 1999). Students rated their psychological safety via a 7-point Likert scale with 1 being "Very inaccurate" and 7 being "Very accurate".

Aggregation

All three variables of interest in this study – psychological safety, relationship conflict, and task conflict – were collected at the individual level. However, as diversity is a team-level metric, it was necessary to aggregate the individual-level variables to the team level to avoid a multilevel mismatch. Within each team, each of these variables of interest were aggregated using

a direct consensus composition model with individual perceptions of each variable as the lower-level construct aggregating to a team-level score on each variable's respective higher-level construct (van Mierlo et al., 2009).

Before aggregation, however, I conducted analyses to justify it. Following recommendations from Woehr et al. (2015) on justifying aggregation, I first calculated ICC(1) and ICC(2) for my dataset of complete cases, and used this as a prerequisite for further consistency analysis. For the three variables of interest (relationship conflict, task conflict, and psychological safety), ICC(1)s ranged from .17-.30 and ICC2s ranged from .45-.64. See Table 2 for ICCs by variable. According to Woehr et al. (2015), ICCs near or above the average estimates reported in the literature (ICC(1): .21, ICC(2): .66) provide support for aggregation. My ICCs were near the means, so I moved forward in my aggregation.

Once ICCs are calculated and considered acceptable, Woehr et al. (2015) then recommend calculating r_{wg} as a measure of within-group agreement on each individual-level variable (see Table 3 for descriptives of r_{wg} by variable). Mean r_{wg} for each variable surpassed the .67 minimum cutoff recommended in Woehr et al. (2015) assuming a slightly skewed distribution, indicating an acceptable level of agreement for aggregation. However, to ensure teams with lower r_{wg} would not skew data, I removed any teams with r_{wg} less than .6 for any of the three variables (24 teams) . Recalculating the ICCs showed slight improvement, with ICC(1)s ranging from .25-.29 and ICC2s ranged from .58-.62 (see Table 4). Because the ICCs were adequate prior to removal of teams with sub-par r_{wg}, I decided to move forward without removing any teams for this reason. Finally, each individual's scores on each variable within each team were averaged to create a team-level variable.

RESULTS

All following analyses used R (R Core Team, 2021) and RStudio (RStudio Team, 2021). See Table 5 for full descriptives and Table 6 for correlations among all study variables.

Hypothesis Testing

To test Hypothesis 1 (regarding the relationship between relationship conflict and psychological safety), I conducted a general linear model regression with psychological safety as the outcome and relationship conflict as the predictor. Results showed that relationship conflict had a significantly negative effect on psychological safety, adjusted $R^2 = 0.29$, $F(1, 253) = 106.6$, $p < .001$. Therefore, Hypothesis 1 was supported. See Table 7 for the full regression table.

A similar test was done to examine Hypothesis 2 (regarding the relationship between task conflict and psychological safety). I conducted a general linear model regression with psychological safety as the outcome and task conflict as the predictor. Results showed that task conflict had a significantly negative effect on psychological safety, adjusted $R^2 = 0.07$, $F(1,253) = 19.92$, $p < .001$. Therefore, Hypothesis 2 was supported. See Table 8 for the full regression table.

To examine Hypothesis 3a, I conducted a moderated linear regression. Like above, psychological safety was set as the outcome and relationship conflict was set as the predictor. To examine gender diversity, Blau_{NS} was included as a moderating term. Neither Blau_{NS} directly nor the moderation between Blau_{NS} and relationship conflict were found to be significant, $p = 0.380$ and $p = 0.478$, respectively. See Table 9 for the full regression table. Gender diversity had no significant effect on the relationship between relationship conflict and psychological safety, not supporting Hypothesis 3a.

To examine Hypothesis 3b, I conducted a similar moderated linear regression to above with psychological safety as the outcome and task conflict as the predictor with $Blau_{NS}$ as a moderating term. Again, Neither $Blau_{NS}$ directly nor the moderation between $Blau_{NS}$ and task conflict were found to be significant, $p = 0.808$ and $p = 0.502$, respectively. Hypothesis 3b was not supported. See Table 10 for the full regression table.

Exploratory Analyses

With the analyses above, gender diversity was not found to have any significant effect on the relationship of either task or relationship conflict and psychological safety. However, it may be that the measure of diversity used – $Blau_{NS}$ – did not adequately address all potential outcomes. Any form of Blau's index only tells the *presence* and *degree* of diversity within a group, but not where that diversity lies. In other words, if Blau's index is low when examining gender diversity in a team, that means there is not a lot of gender diversity present; however, this does not indicate if that means the team was majority men or women, just that there was a majority of some kind.

Circling back to the literature above, there is evidence that men and women may have differing conflict management styles, with men being more dominant and women being more cooperative (e.g., Brewer et al., 2002; Rahim & Katz, 2019). Perhaps, then, it is not the presence or balance of these two conflict management styles that lessens the effect of conflict, but rather the presence of that specific cooperative conflict management style. In this instance, it may be that the proportion of women is directly related to how strongly conflict affects psychological safety, rather than gender diversity generally.

To examine this, I calculated the proportion of women in each team. There was an average of 51% women on each team, with a standard deviation of 31%. I then ran a series of

moderated linear regressions similar to those used to test Hypotheses 3a and 3b above. For the first model, my outcome was psychological safety and my antecedent was relationship conflict moderated by proportion of women on each team. Direct effects of proportion of women were not significant ($p = 0.321$), nor was the moderating effect of proportion of women on the relationship between relationship conflict and psychological safety ($p = 0.361$).

Next, I tested this same relationship with task conflict rather than relationship conflict. In this model, the direct effects of proportion of women were significant ($p = 0.021$), as was the moderating effect of proportion of women on the relationship between task conflict and psychological safety, $p = .005$. Comparing change in adjusted R^2, however, showed that this moderating term had a small impact on proportion of variance accounted for in the model ($\Delta R^2 = .03$). See Appendix B for full regression tables for both models.

To probe this interaction, I used the probe_interaction function within the "interactions" package (Long, 2019) in R Studio. Simple slopes showed that, counter to theorizing above, teams with a higher proportion of women had a *stronger* relationship between psychological safety and task conflict; in other words, the more women on a team, the more task conflict negatively influenced psychological safety.

DISCUSSION

The first contribution of this study was to add to literature on the direct effects of conflict on psychological safety. Specifically, as mentioned above, most studies on psychological safety tend to examine it from the individual-level rather than the group-level (Frazier et al., 2017). I examined psychological safety on the group-level, contributing to this less-saturated sector of psychological safety literature.

Both task and relationship conflict were found to have significant negative effects on psychological safety, supporting Hypotheses 1 and 2. This follows theorizing and previous literature positing the negative effect of conflict on psychological safety. This study is, to my knowledge, only the second study directly examining the link between these constructs (with the first being Wilkens & London, 2006). Meta-analyses such as Frazier et al. (2017) have historically not examined conflict in relation to psychological safety due to the limited empirical research on the topic. Through this paper, then, I attempt to fill that hole, adding another piece of evidence supporting the negative relationship between conflict and psychological safety. Further research should continue to examine this relationship.

The second major contribution of this study was examining gender diversity as a contextual moderator rather than simply as a predictor. Mixed literature on the topic indicates that effects of gender diversity aren't simply positive or negative; more nuance is required to truly evaluate how gender diversity affects team processes. I contribute to the study of gender diversity by highlighting an alternative perspective that may be employed to uncover these nuanced relationships.

However, the relationships hypothesized in Hypotheses 3a and 3b were not found. One potential explanation of this was that it is having women on a team, not just gender diversity, that

lessens the negative effects of conflict; this explanation was also not supported, as shown through the exploratory analyses. Rather, through these exploratory analyses, proportion of women on a team was found to *exacerbate* the negative effects of task conflict on psychological safety.

The explanation of these findings may come not from conflict management literature, but rather from literature on the backlash effect (Rudman, 1998). The backlash effect finds that individuals violating prescriptive norms may be subject to negative judgments and outcomes from others (Rudman, 1998), potentially undermining their psychological safety. A common example of this occurs when women – who are prescribed to adhere to communal norms such as warmth – act more agentically, e.g., act with dominance in a group setting (Heilman, 2001).

In the workplace, women may be more likely to act in this way when they are in a male-dominated organization (Rudman & Phelan, 2008); similarly, in the context of students, perhaps women are more likely to act agentically and counterstereotypically when they are in a male-dominated major, exacerbating backlash effects in these settings. STEM fields are traditionally male-dominated, so it would stand to reason that women in these majors may act more agentically. In this study's sample, 75% of students were in a STEM or STEM-adjacent major (26% in Business and 49% in STEM), and women comprised 40% and 34% of those students, respectively. Therefore, women in this study may have experienced backlash or been wary of experiencing backlash due to their role-incongruent behavior (acting agentically).

It is also important to note that proportion of women on a team was significantly and negatively related to task conflict, $r = -.24$, $p < .01$. Having more women on a team decreases the likelihood of task conflict on that team occurring, which corresponds with the discussion of conflict management styles above. However, due to so many women being in a male-dominated

major and being used to experiencing backlash because of their agentic behavior, women may have been on-edge during their team interactions. When task conflict did occur, women felt unsafe in the team.

Exacerbating this effect, there is literature that finds that women may exhibit as much or even more backlash toward other women acting unstereotypically than men would. For instance, Marques (1990) discussed the black sheep effect, in which norm violations by an individual are judged more harshly by their own ingroup than by an outgroup. Similarly, Bargh (1989) discussed how women react negatively to power displays by other women automatically, engaging in a form of unconscious self-oppression. This effect may be even more pronounced when a woman believes they rely on another woman for their success (Rudman, 1998), as they may in a team. This twist on backlash in teams dominated by women is another major yet unexpected contribution of this study. Literature has examined this effect in a hiring context, but not often in a team-based context and never with psychological safety as the outcome.

Why, however, was this moderating effect found for only task conflict, and not relationship conflict? The answer may lie in Tajfel's (1981) Social Identity Theory, as discussed briefly in the literature review above. Social Identity Theory states that individuals identify with others who share similarities with them, forming an ingroup. Attitudes toward those within one's ingroup tend to be more favorable than those not in one's ingroup. In the context of this study, women in a team may have formed a sort of ingroup due to their shared characteristic of gender. As relationship conflict has to do with liking individuals within your team, women in these ingroups may have subjectively liked each other (i.e., relationship conflict would not be exacerbated). However, the backlash affects discussed above would still apply to task conflict,

which does not have to do with subjective liking; norm violations due to task-related behavior would be judged harshly by these ingroup women.

Limitations and Future Directions

This study did have a few limitations. Firstly, while the dataset used was vast, when filtered for the specifications the study required, it became substantially smaller (n = 255 teams). However, I used G*Power (Faul et al., 2007) to conduct a post-hoc power analysis. Results showed I achieved a power of 1.00 with an effect size of $f^2 = .561$. This indicates that power was not a concern for this study.

Another aspect of the data that may have limited this study was the use of teams consisting of students, who are usually younger than working professionals. We argue these teams are in fact true project teams as the students worked together for an extended period of time on a task that influenced their grade. However, as the participants were college students, their average age was 22.08 (SD = 5.25). The students' youth may have affected their willingness to take their team project at hand and/or the peer ratings seriously, potentially shrouding significant effects. Additionally, 75% of participants in this study were either STEM or Business majors; as this study was at the team-level, I cannot control for individual students' majors, but nevertheless this majority should be considered when interpreting results.

Future research should potentially examine the differential effect of these teams being comprised of students versus working professionals, and also consider various factors of team composition (e.g., age). This, however, may be a small piece of the larger issue with research being dominated by student samples due to convenience; generally, researchers should strive to compare their research with student samples to that with samples of more relevant populations.

In addition, the backlash effect of women on other women in teams found in this paper should be examined more thoroughly. It seems this effect may exacerbate the negative effects of conflict on psychological safety in certain situations; however, other outcomes like cohesion and satisfaction are important to examine to get a full picture of this effect. Future researchers should continue examining the backlash effect in various arenas and with various moderators to better understand this complex effect.

In this study, I examined one factor of diversity: gender. I examined the gender breakdown of a team as a contextual moderator of the relationship between various forms of conflict and team psychological safety. However, when examining diversity, it is important to consider intersectional effects (Crenshaw, 2017). For instance, one's race may interact with one's gender to exacerbate backlash received, or perhaps mitigate it and other negative effects. For the present study, race diversity (i.e., white vs. non-white) did not correlate significantly with any variables examined; it is therefore reasonable to expect race would not have any significant effects if added into this study's models. However, future studies should strive to examine multiple forms of diversity, separate and together, to fully understand the full picture of diversity and uncover potentially complex interactions.

In summary, I examined the relationships between task and relationship conflict on psychological safety with gender diversity as a contextual moderator. While literature supported hypotheses predicting that gender diversity would significantly lessen the negative effects of conflict on psychological safety, I did not find support for this in my analyses. Gender diversity had no significant effect on psychological safety directly nor as a moderator of the relationship between either task or relationship conflict and psychological safety. Only during exploratory

analyses using proportion of women on the team rather than gender diversity generally did I find any evidence of significant moderation, and it was in the opposite direction than predicted.

Though this one significant moderation was an interesting finding, the lack of significance otherwise (and the overall small effect sizes) indicate that gender diversity may not have a positive or negative effect on this set of relationships; instead, gender diversity may just not affect the relationship between conflict and psychological safety. This in itself is an interesting finding; null results in organizational science research are often ignored or underutilized but may contribute to a body of literature in a significant way if considered. In the body of diversity literature, if null findings are compiled they may at least indicate in which situations it is important to consider diversity further. It seems, in this case, gender diversity alone does not have a significant effect on the relationship between team conflict and psychological safety. Instead, only proportion of women (rather than a more general diversity index) had a significant moderating effect. If replicated, this result may indicate that gender diversity alone is not a significant factor to consider for the relationship between team conflict and psychological safety, and instead we should examine the proportion of women on a team (or other more specific diversity measures) to uncover nuanced relationships.

REFERENCES

Acker, J. (1990). Hierarchies, jobs, bodies: A theory of gendered organizations. *Gender & Society*, *4*(2), 139–158. https://doi.org/10.1177/089124390004002002

Anderson, K. J., & Leaper, C. (1998). Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how. *Sex Roles*, *39*(3/4), 225–252. https://doi.org/10.1023/A:1018802521676

Auh, S., Spyropoulou, S., Menguc, B., & Uslu, A. (2014). When and how does sales team conflict affect sales team performance? *Journal of the Academy of Marketing Science*, *42*(6), 658–679. https://doi.org/10.1007/s11747-014-0368-7

Barclay, P. (2010). Altruism as a courtship display: Some effects of third-party generosity on audience perceptions. *British Journal of Psychology*, *101*(1), 123–135. https://doi.org/10.1348/000712609X435733

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. *Unintended Thought.*, 3–51.

Bear, J. B., & Woolley, A. W. (2011). The role of gender in team collaboration and performance. *Interdisciplinary Science Reviews*, *36*(2), 146–153. https://doi.org/10.1179/030801811X13013181961473

Bendersky, C., & Hays, N. A. (2012). Status conflict in groups. *Organization Science*, *23*(2), 323–340. https://doi.org/10.1287/orsc.1110.0734

Berger, J., Cohen, B. P., & Zelditch, M. (1972). Status characteristics and social interaction. *American Sociological Review*, *37*(3), 241. https://doi.org/10.2307/2093465

Biemann, T., & Kearney, E. (2010). Size does matter: How varying group sizes in a sample affect the most common measures of group diversity. *Organizational Research Methods*, *13*(3), 582–599. https://doi.org/10.1177/1094428109338875

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. Free Press.

Bono, J. E., Boles, T. L., Judge, T. A., & Lauver, K. J. (2002). The role of personality in task and relationship conflict. *Journal of Personality*, *70*(3), 311–344. https://doi.org/10.1111/1467-6494.05007

Brescoll, V. L. (2011). Who takes the floor and why: Gender, power, and volubility in organizations. *Administrative Science Quarterly*, *56*(4), 622–641. https://doi.org/10.1177/0001839212439994

Brewer, N., Mitchell, P., & Weber, N. (2002). Gender role, organizational status, and conflict management styles. *International Journal of Conflict Management*, *13*(1), 78–94. https://doi.org/10.1108/eb022868

Caudron, S. (1998, September). Keeping team conflict alive. *Training & Development*, *52*(9), 48+. Gale Academic OneFile.

Costa, P. L., Passos, A. M., & Bakker, A. B. (2015). Direct and contextual influence of team conflict on team resources, team work engagement, and team performance. *Negotiation and Conflict Management Research*, *8*(4), 211–227. https://doi.org/10.1111/ncmr.12061

Crenshaw, K. (2017). *On intersectionality: Essential writings*. New Press.

De Dreu, C. K. W., & Van De Vliert, E. (Eds.). (1997). *Using conflict in organizations*. SAGE.

De Dreu, C. K. W., & Van Vianen, A. E. M. (2001). Managing relationship conflict and the effectiveness of organizational teams. *Journal of Organizational Behavior*, *22*(3), 309–328. https://doi.org/10.1002/job.71

De Dreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology*, *88*(4), 741–749. https://doi.org/10.1037/0021-9010.88.4.741

de Wit, F. R. C., Jehn, K. A., & Scheepers, D. (2013). Task conflict, information processing, and decision-making: The damaging effect of relationship conflict. *Organizational Behavior and Human Decision Processes*, *122*(2), 177–189. https://doi.org/10.1016/j.obhdp.2013.07.002

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Farrelly, D. (2011). Cooperation as a signal of genetic or phenotypic quality in female mate choice? Evidence from preferences across the menstrual cycle: Ovulation and cooperation in mate choice. *British Journal of Psychology*, *102*(3), 406–430. https://doi.org/10.1348/000712610X532896

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fenwick, G. D., & Neal, D. J. (2001). Effect of gender composition on group performance. *Gender, Work & Organization*, *8*(2), 205–225. https://doi.org/10.1111/1468-0432.00129

Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2017). Psychological safety: A meta-analytic review and extension. *Personnel Psychology*, *70*(1), 113–165. https://doi.org/10.1111/peps.12183

Friedman, R. A., Tidd, S. T., Currall, S. C., & Tsai, J. C. (2000). What goes around comes around: The impact of personal conflict style on work conflict and stress. *International Journal of Conflict Management*, *11*(1), 32–55. https://doi.org/10.1108/eb022834

Gamero, N., González-Romá, V., & Peiró, J. M. (2008). The influence of intra-team conflict on work teams' affective climate: A longitudinal study. *Journal of Occupational and Organizational Psychology*, *81*(1), 47–69. https://doi.org/10.1348/096317907X180441

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*(4), 657–674. https://doi.org/10.1111/0022-4537.00234

Heilman, M. E., & Welle, B. (2006). Disadvantaged by diversity? The effects of diversity goals on competence perceptions. *Journal of Applied Social Psychology*, *36*(5), 1291–1319. https://doi.org/10.1111/j.0021-9029.2006.00043.x

Hirschfeld, R. R., Jordan, M. H., Feild, H. S., Giles, W. F., & Armenakis, A. A. (2005). Teams' female representation and perceived potency as inputs to team outcomes in a predominantly male field setting. *Personnel Psychology*, *58*(4), 893–924. https://doi.org/10.1111/j.1744-6570.2005.00892.x

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Janssen, O., Van De Vliert, E., & Veenstra, C. (1999). How task and person conflict shape the role of positive interdependence in management teams. *Journal of Management*, *25*(2), 117–141. https://doi.org/10.1177/014920639902500201

Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, *40*(2), 256. https://doi.org/10.2307/2393638

Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, *44*(2), 238–251. https://doi.org/10.2307/3069453

Kessel, M., Kratzer, J., & Schultz, C. (2012). Psychological safety, knowledge sharing, and creative performance in healthcare teams. *Creativity and Innovation Management*, *21*(2), 147–157. https://doi.org/10.1111/j.1467-8691.2012.00635.x

Knoppers, A., Meyer, B. B., Ewing, M., & Forrest, L. (1991). Opportunity and work behavior in college coaching. *Journal of Sport and Social Issues*, *15*(1), 1–20. https://doi.org/10.1177/019372359101500101

Kostopoulos, K. C., & Bozionelos, N. (2011). Team exploratory and exploitative learning: Psychological safety, task conflict, and team performance. *Group & Organization Management*, *36*(3), 385–415. https://doi.org/10.1177/1059601111405985

Lau, R. S., & Cobb, A. T. (2010). Understanding the connections between relationship conflict and performance: The intervening roles of trust and exchange. *Journal of Organizational Behavior*, *31*(6), 898–917. https://doi.org/10.1002/job.674

Lee, H. W., Choi, J. N., & Kim, S. (2018). Does gender diversity help teams constructively manage status conflict? An evolutionary perspective of status conflict, team

psychological safety, and team creativity. *Organizational Behavior and Human Decision Processes*, *144*, 187–199. https://doi.org/10.1016/j.obhdp.2017.09.005

Long, J. A. (2019). *interactions: Comprehensive, user-friendly toolkit for probing interactions*. https://cran.r-project.org/package=interactions

Loughry, M. L., Ohland, M. W., & DeWayne Moore, D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement*, *67*(3), 505–524. https://doi.org/10.1177/0013164406292085

Marques, J. M. (1990). The black-sheep effect: Out-group homogeneity in social comparison settings. In D. Abrams & M. A. Hogg (Eds.), *Social identity theory: Constructive and critical advances* (pp. 131–151).

McLeod, P. L., Lobel, S. A., & Cox, T. H. (1996). Ethnic diversity and creativity in small groups. *Small Group Research*, *27*(2), 248–264. https://doi.org/10.1177/1046496496272003

Medina, F. J., Munduate, L., Dorado, M. A., Martínez, I., & Guerra, J. M. (2005). Types of intragroup conflict and affective reactions. *Journal of Managerial Psychology*, *20*(3/4), 219–230. https://doi.org/10.1108/02683940510589019

Myaskovsky, L., Unikel, E., & Dew, M. A. (2005). Effects of gender diversity on performance and interpersonal behavior in small work groups. *Sex Roles*, *52*(9–10), 645–657. https://doi.org/10.1007/s11199-005-3732-8

Ortmann, A., & Tichy, L. K. (1999). Gender differences in the laboratory: Evidence from prisoner's dilemma games. *Journal of Economic Behavior & Organization*, *39*(3), 327–339. https://doi.org/10.1016/S0167-2681(99)00038-4

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rahim, M. A., & Katz, J. P. (2019). Forty years of conflict: The effects of gender and generation on conflict-management strategies. *International Journal of Conflict Management*, *31*(1), 1–16. https://doi.org/10.1108/IJCMA-03-2019-0045

Ridgeway, C. (2014). Why status matters for inequality. *American Sociological Review*, *79*(1), 1–16. https://doi.org/10.1177/0003122413515997

RStudio Team. (2021). *RStudio: Integrated development environment for R*. RStudio, PBC. http://www.rstudio.com/

Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, *74*(3), 629–645.

Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior*, *28*, 61–79. https://doi.org/10.1016/j.riob.2008.04.003

Salas, E., Burke, C. S., & Cannon-Bowers, J. A. (2000). Teamwork: Emerging principles. *International Journal of Management Reviews*, *2*(4), 339–356. https://doi.org/10.1111/1468-2370.00046

Savage, S. V., Dippong, J., & Melamed, D. (2020). Status and competitive choice. *Social Science Research*, *88–89*, 102430. https://doi.org/10.1016/j.ssresearch.2020.102430

Simons, T. L., & Peterson, R. S. (2000). Task conflict and relationship conflict in top management teams: The pivotal role of intragroup trust. *Journal of Applied Psychology*, *85*(1), 102–111. https://doi.org/10.1037/0021-9010.85.1.102

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge University Press.

van den Oever, K., & Beerens, B. (2021). Does task-related conflict mediate the board gender diversity–organizational performance relationship? *European Management Journal*, *39*(4), 445–455. https://doi.org/10.1016/j.emj.2020.09.008

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing group-level constructs from individual-level survey data. *Organizational Research Methods*, *12*(2), 368–392. https://doi.org/10.1177/1094428107309322

Wageman, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly*, *40*(1), 145. https://doi.org/10.2307/2393703

Wall, J. A., & Callister, R. R. (1995). Conflict and its management. *Journal of Management*, *21*(3), 515–558.

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Wilkens, R., & London, M. (2006). Relationships between climate, process, and performance in continuous quality improvement groups. *Journal of Vocational Behavior*, *69*(3), 510–523. https://doi.org/10.1016/j.jvb.2006.05.005

Williams, M., & Polman, E. (2015). Is it me or her? How gender composition evokes interpersonally sensitive behavior on collaborative cross-boundary projects. *Organization Science*, *26*(2), 334–355. https://doi.org/10.1287/orsc.2014.0941

Woehr, D. J., Loignon, A. C., Schmidt, P. B., Loughry, M. L., & Ohland, M. W. (2015). Justifying aggregation with consensus-based constructs: A review and examination of

cutoff values for common aggregation indices. *Organizational Research Methods*, *18*(4), 704–737. https://doi.org/10.1177/1094428115582090

Table 1

Means, Standard Deviations, and Correlations Among Calculated Blau Indices

Variable	*M*	*SD*	1	2
1. Blau	0.28	0.22		
2. $Blau_N$	0.37	0.29	.99**	
3. $Blau_{NS}$	0.56	0.43	.99**	1.00**

Note. M and *SD* are used to represent mean and standard deviation, respectively. ** indicates $p < .01$.

Table 2

ICC(1) and ICC(2) by Variable

Variable	ICC(1)	ICC(2)
Relationship conflict	0.30	0.64
Task conflict	0.24	0.56
Psychological safety	0.17	0.45

Note. $N = 1030$ individuals nested in 255 teams for this analysis.

Table 3

Descriptive Statistics of r_{wg} by Variable

Variable	*M*	*SD*	Min	Max
Relationship conflict r_{wg}	0.93	0.13	0.09	1.00
Task conflict r_{wg}	0.89	0.11	0.33	1.00
Psychological safety r_{wg}	0.88	0.14	0.00	1.00

Note. All variables above were aggregated to the team-level. $N = 1030$ individuals nested in 255 teams.

Table 4

ICC(1) and ICC(2) by Variable with Teams with r_{wg} < .6 Removed

Variable	ICC(1)	ICC(2)
Relationship conflict	0.25	0.58
Task conflict	0.29	0.62
Psychological safety	0.28	0.61

Note. N = 790 individuals nested in 231 teams for this analysis.

Table 5

Descriptive Statistics of Dataset Filtered Only for Complete Cases

Variable	*N*	*M*	*SD*	Min	Max
Team size	255	4.45	1.03	3.00	7.00
Percent female	255	0.51	0.31	0.00	1.00
Psychological safety	255	5.97	0.48	4.48	7.00
Relationship conflict	255	1.20	0.30	1.00	2.78
Task conflict	255	1.48	0.37	1.00	2.67
$Blau_{NS}$	255	0.61	0.41	0.00	1.00

Note. All variables above aggregated to the team-level.

Table 6

Means, Standard Deviations, and Pair-wise Correlations of All Variables

Variable	*M*	*SD*	1	2	3	4	5	6
1. Team size	4.45	1.03						
2. Percent women	0.51	0.31	.15*					
3. Percent white†	0.65	0.26	-.07	-.12				
4. Relationship conflict	1.20	0.30	.06	.01	-.04	(.81)		
5. Task conflict	1.48	0.37	.16*	-.24**	.00	.54**	(.80)	
6. Psychological safety	5.97	0.48	-.04	-.03	.10	-.54**	-.27**	(.71)
7. $Blau_{NS}$	0.31	0.20	-.06	.18**	-.01	.04	-.11	-.07

Note. M and *SD* are used to represent mean and standard deviation, respectively. $N = 255$ for each pairwise correlation. * indicates $p < .05$. ** indicates $p < .01$. Values on the diagonal indicate Cronbach's alpha for each variable of interest.
† Percent white was calculated within each team as a supplementary demographic variable but was not used in any analyses.

Table 7

Regression Results for Hypothesis 1 (Relationship Conflict on Psychological Safety)

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	*df*	Effect Size
(Intercept)	7.00**	[6.80, 7.21]							
Relationship Conflict	-0.86**	[-1.03, -0.70]	-0.54	[-0.65, -0.44]	.30	[.21, .38]	-.54**	1, 253	
									R^2 = .296**
									95% CI[.21,.38]

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. *df* indicates degrees of freedom. ΔR^2 indicates the change in R^2 when adding the moderation to the model in a step-wise fashion.
* indicates $p < .05$. ** indicates $p < .01$.

Table 8

Regression Results for Hypothesis 2 (Task Conflict on Psychological Safety)

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	*df*	Effect Size
(Intercept)	6.50**	[6.26, 6.73]							
Task Conflict	-0.36**	[-0.51, -0.20]	-0.27	[-0.39, -0.15]	.07	[.02, .14]	-.27**		
								1, 253	R^2 = .073**
									95% CI[.02,.14]

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. *df* indicates degrees of freedom. ΔR^2 indicates the change in R^2 when adding the moderation to the model in a step-wise fashion.
* indicates $p < .05$. ** indicates $p < .01$.

Table 9

Regression Results for Hypothesis 3A (Relationship Conflict on Psychological Safety Moderated by $Blau_{NS}$)

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*df*	Effect Size	Δ R^2
(Intercept)	5.97**	[5.92, 6.02]					
Relationship Conflict	-0.87**	[-1.04, -0.70]	.29	[.19, .38]			
$Blau_{NS}$	-0.05	[-0.18, 0.07]	.00	[-.01, .01]			
Relationship Conflict x $Blau_{NS}$	0.19	[-0.34, 0.71]	.00	[-.01, .01]			.002
					3, 251	R^2 = .300** 95% CI[.20,.38]	

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. *df* indicates degrees of freedom. Δ R^2 indicates the change in R^2 when adding the moderation to the model in a step-wise fashion.

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

Table 10

Regression Results for Hypothesis 3B (Task Conflict on Psychological Safety Moderated by $Blau_{NS}$)

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*df*	Effect Size	Δ R^2
(Intercept)	5.97**	[5.91, 6.03]					
Task Conflict	-0.38**	[-0.54, -0.22]	.08	[.02, .14]			
$Blau_{NS}$	-0.12	[-0.26, 0.02]	.01	[-.01, .03]			
Task Conflict x $Blau_{NS}$	-0.13	[-0.52, 0.26]	.00	[-.01, .01]			.006
					3, 251	R^2 = .086** 95% CI[.02,.15]	

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. *df* indicates degrees of freedom. Δ R^2 indicates the change in R^2 when adding the moderation to the model in a step-wise fashion.

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

APPENDIX A: Information on Dataset Before Filtering for Complete Cases

Table 1A

Descriptive Statistics of Dataset by Team

Variable	*n*	*M*	*SD*	Min	Max
Team size	4,642	4.26	0.98	3.00	10.00
Percent female	4,642	0.37	0.30	0.00	1.00
Psychological safety	326	5.99	0.47	4.48	7.00
Relationship conflict	1,495	1.23	0.35	1.00	4.33
Task conflict	1,495	1.54	0.40	1.00	4.25
$Blau_{NS}$	4,642	0.56	0.43	0.00	1.00

Note. All variables above were aggregated to the team-level.

APPENDIX B: Regression Tables for Exploratory Analyses

Table 1B

Regression Results for Exploratory Analysis 1 (Relationship Conflict on Psychological Safety Moderated by Percent Women)

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*df*	Effect Size	Δ R^2
(Intercept)	5.97**	[5.92, 6.02]					
Relationship Conflict	-0.89**	[-1.07, -0.71]	.28	[.18, .37]			
Percent women	-0.04	[-0.20, 0.13]	.00	[0.00, .01]			
Relationship Conflict x Percent women	0.30	[-0.35, 0.95]	.00	[-.01, .01]			.003
					3, 251	R^2 = .299** 95% CI[.20,.38]	

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. *df* indicates degrees of freedom. Δ R^2 indicates the change in R^2 when adding the moderation to the model in a step-wise fashion.
* indicates $p < .05$. ** indicates $p < .01$.

Table 2B

Regression Results for Exploratory Analysis 2 (Task Conflict on Psychological Safety Moderated by Percent Women)

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*df*	Effect Size	Δ R^2
(Intercept)	5.95**	[5.89, 6.01]					
Task Conflict	-0.42**	[-0.58, -0.26]	.09	[.03, .16]			
Percent women	-0.14	[-0.33, 0.05]	.01	[-.01, .03]			
Task Conflict x Percent women	-0.70**	[-1.19, -0.21]	.03	[-.01, .07]			.032
					3, 251	R^2 = .112** 95% CI[.04,.18]	

Note. A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. sr^2 represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. *df* indicates degrees of freedom. Δ R^2 indicates the change in R^2 when adding the moderation to the model in a step-wise fashion.

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