(Mitigating) the self-fulfillment of gender stereotypes in teams
Meyer, Bertolt; Dijk, Hans van; Engen, Marloes van

Published in:
Journal of Applied Psychology

DOI:
10.1037/apl0000995

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 14. Sep. 2023
We challenge the social categorization perspective in the team diversity literature by arguing that stereotypes and not favoritism for members of the same social category govern processes and dynamics in gender-diverse teams. We posit that team members’ gender and task stereotypes generate competence attributions that shape individual team members’ dominance behavior and performance in a self-fulfilling way: Team members who are attributed more competence behave more dominantly and outperform those who are attributed less competence. We further argue that pro-diversity beliefs may prevent this self-fulfilling tendency of stereotypes by inhibiting individuals’ stereotype-confirming behavior. Hypotheses were tested with 97 gender-heterogeneous four-person student teams working on stereotypically masculine- or feminine-typed problems. Team members estimated each other’s competence prior to collaboration. Diversity beliefs were manipulated to be either pro-diversity or pro-similarity and dominance was observed with behavioral coding. Multilevel path modeling showed that competence attributions mediated the effects of stereotypical gender-task fit on individual dominance behavior and performance under pro-similarity beliefs but not under pro-diversity beliefs. Our study thus shows that the self-fulfilling tendencies of gender stereotypes in teams can be mitigated by instituting pro-diversity beliefs.

**Keywords:** stereotyping, teams, gender, diversity, diversity beliefs

Similarly, research on individual team members’ behavior and individual performance in the context of diverse teams suggests that being different influences interpersonal dynamics, such that a target member who is perceived as different from the rest of the team will receive less support and experience more conflict, which is likely to harm the target member’s display of competence and individual performance (Guillaume et al., 2014; Meyer, 2017; Shemla et al., 2016).

Interestingly, beyond the team context, research on gender in organizations primarily relies on stereotyping theory to explain and predict how gender differences shape attributions, interpersonal behaviors, and individual performance (Cuddy et al., 2011; Koch et al., 2015). For example, research on the lack-of-fit model (Heilman, 1983, 2012) and role congruence theory (Eagly & Karau, 2002; Eagly & Sczesny, 2019) shows that stereotypes related to gendered tasks and jobs influence how individuals perceive themselves and each other, such that higher levels of congruence between task and gender stereotypes result in higher attributions of competence. In turn, research on stereotype threat (Schmader et al., 2008; Steele & Aronson, 1995), stereotype lift (Walton & Cohen, 2003), and stereotype susceptibility (Shih et al., 1999) suggest that such stereotype-based competence attributions shape individual performance in a self-fulfilling way: Low levels of task-gender stereotype congruence tend to result in low levels of individual performance, whereas high levels of task-gender stereotype congruence tend to result in high levels of individual performance.

In light of these different explanations for the consequences of gender diversity, in the current article, we aim to understand whether...
interpersonal attributions, behaviors, and individual performance in gender-diverse teams are primarily influenced by being different (as proposed by social categorization theory) or by stereotypes. We build on research on the function of stereotypes (Fiske et al., 2002; Fiske & Neuberg, 1990; van Dijk et al., 2017) to argue that in teams, members frequently need to determine who is the most competent (Barton & Bunderson, 2014; Wittenbaum & Bowman, 2005), and that stereotypes provide an implicit understanding of who may be the most competent team member in a given situation. In turn, these competence attributions shape interpersonal behavior and individual performance in the team, such that perceiving members enable a target team member to display more dominant behavior and achieve higher levels of individual performance when the target member is considered more competent at a task at hand.

Despite arguing that stereotypes govern attributions and interpersonal dynamics in gender-diverse teams and can be functional, we acknowledge the potential detrimental effects that stereotypes can have in gender-diverse teams: By shaping interpersonal behavior and individual performance in self-fulfilling ways, stereotypes reinforce and legitimize the gender segmentation of tasks and jobs (Jost & Banaji, 1994; van Dijk & van Engen, 2019). We, therefore, pursue research that examines how the self-fulfilling tendencies of gender stereotypes can be prevented. By combining insights from research from the areas of stereotyping and team diversity, we argue that making members aware of the value of diverse perspectives in teamwork may be an effective intervention to overcome the detrimental effects of stereotypes (Bezrukova et al., 2016; Kalinski et al., 2013). Specifically, whereas stereotype-based attributions of competence in teams may serve as indicators of assumed competence of team members and trigger behavior that confirms such expectations (Fiske et al., 2002; van Dijk et al., 2017), we assert that pro-diversity beliefs challenge such assumptions by making team members aware that all members of diverse teams can provide meaningful contributions. As such, we argue that pro-diversity beliefs stimulate individuals not to rely on stereotype-based competition attributions in team contexts, and instead be open to and appreciative of the input of each individual member.

Our study provides three main contributions. First, we contribute to theory and research on diversity in teams by challenging the long-held assumption that group processes related to in- and outgroup processes drive behavior and performance in diverse teams (e.g., Homan et al., 2020; van Knippenberg et al., 2004). We do this by developing a cross-level theory on how stereotypes, and not social group membership, are the primary drivers of interpersonal attributions, behavior, and individual performance in gender-diverse teams. Second, we extend theory and research on stereotyping in organizations (Cuddy et al., 2011; Eagly & Sczesny, 2019; Heilman & Caleo, 2018) by showing that stereotypes not only affect outcomes of those stereotyped, but also influence dynamics in diverse teams in a stereotype-reinforcing fashion. Third, by combining insights from diversity research and stereotyping research into one unified cross-level theory, we contribute to research on how the reinforcing tendencies of gender stereotypes can be mitigated. We provide these contributions in a lab study in which we examine competence attributions made by team members about individuals in the team and its effect on the target individuals’ dominance behavior and individual performance on a stereotypically masculine or a stereotypically feminine task.

Disconnections Between Theory and Research on Diversity in Teams and Stereotypes in Organizations

There is a long and rich history of theory and research arguing and showing the pervasive effects of stereotyping in organizations (e.g., Biernat & Sesko, 2018; Eagly et al., 1992; Heilman et al., 2004; Koch et al., 2015; Thomas-Hunt & Phillips, 2004). Unfortunately, those insights have been largely absent in theories on how diversity shapes behavior and performance in teams (van Dijk et al., 2017). For decades, research on work group diversity and relational demography has been predominated by a bitheoretical approach consisting of the information/decision-making and the social categorization perspectives (see Meyer, 2017; Shenla et al., 2016; van Dijk et al., 2017; van Knippenberg & Schippers, 2007; Williams & O’Reilly, 1998, for reviews).

The information/decision-making perspective proposes that differences among team members on any attribute (such as gender, ethnicity, educational background, and functional background) can enhance individual-level and team-level information processing because team members from different social categories have different knowledge, experiences, and views (van Knippenberg et al., 2004). On the other hand, the social categorization perspective suggests that team members’ potential negative reactions to diversity threaten this informational benefit of diversity. If team members identify more strongly with team members from their own social group (i.e., the ingroup) compared to other team members or the team as a whole, subgroups form such that members of the same team perceive each other as belonging to different social groups (Carton & Cummings, 2012; Harrison & Klein, 2007). Such subgroups give rise to ingroup bias, causing members to prefer members of their own subgroup over other team members. This ingroup bias among team members subsequently is believed to harm cohesion and information processing, and result in conflicts among team members and decreased individual and team performance (Meyer, 2017; van Knippenberg et al., 2004; see also research on faultlines and subgroups, e.g., Thatcher & Patel, 2012).

The social categorization perspective thus suggests that dissimilarity, once salient, can only yield negative consequences because perceived differences mean perceiving the other team members as an outgroup, which results in a devaluation of the different colleague. From the perspective of a single woman member of an otherwise all-men four-person team, this entails that the social categorization perspective predicts that men will perceive her as different from them, resulting in an unfavorable perception of her by her male colleagues, which results in less trust and exchange between her and her peers. Given the predominance of this social categorization perspective in theory and research on team diversity, salient differences between team members are seen as something that is problematic (e.g., Harrison & Klein, 2007; Srikanth et al., 2016; van Dijk et al., 2012; van Knippenberg et al., 2004; Williams & O’Reilly, 1998). Aside from some occasional references that negative stereotypes may strengthen social categorization and its negative outcomes (e.g., Guillaume et al., 2017; Joshi & Roh, 2009), stereotypes do not play a role in studies that are grounded in this bitheoretical approach.

Interestingly, this assumption of ingroup favoritism from the social categorization perspective is at odds with stereotype theory and research, which suggests that stereotypes can cause individuals to hold more favorable attributions toward outgroup compared to ingroup members (Fiske et al., 2002; Jost et al., 2004;
Rudman & Fairchild, 2004). Stereotypes are defined as “cognitive structures that provide knowledge, beliefs, and expectations about individuals based on their social group membership” (Quattrifogli & Macrae, 2011, pp. 216–217), and one of its main functions is forming an impression of a target’s competence (Abele et al., 2021; Fiske et al., 2002). In particular, in work contexts, such impressions of competence can be useful. For example, selecting a job candidate, assigning a person for a task, determining which suggestion or idea to follow, and performing job reviews are all to a considerable extent, based on impressions of competence (Cuddy et al., 2011).

Despite the potential functionality of stereotypes, most research on stereotyping in organizations tends to focus on its negative consequences by building on Heilman’s (1983, 2012) lack-of-fit model and Eagly and Karau’s (2002) role congruity theory. These are conceptually similar theories that explain that (wo)men are often attributed lower levels of competence for masculine (feminine) typed tasks because stereotypes pertaining to what competencies are needed to perform well on a task are incongruent with gender stereotypes. For example, studies have shown that the stereotypes of leaders and of women are incongruent (lack-of-fit), and women are less likely to be selected into leadership positions and once selected experience backlash for acting as a leader (Eagly & Karau, 2002), despite women tending to be more effective leaders (for meta-analyses, see Eagly et al., 2003; Koenig et al., 2011).

Notwithstanding the importance of mapping the negative consequences of inaccurate stereotypes and attempts at remedy, the sole focus on the negative consequences of stereotyping in organizations results in two limitations. First, it obscures an understanding of the functions of stereotyping. For example, there are many more studies on how a lack of task-gender congruence harms target’s individual performance (cf. research on stereotype threat; Steele & Aronson, 1995) than the presence of task-gender congruence benefits individual performance (cf. research on stereotype lift and stereotype susceptibility; Appel & Weber, 2021; Walton & Cohen, 2003). Second, in examining the negative consequences of stereotyping for target individuals, studies tend to focus on the individual level without taking the interpersonal, team, and organizational context in which they are embedded into account. As a consequence, cross-level studies on how interpersonal relationships in a team context shape or influence the presence and consequences of task-gender congruence are generally absent.

To address these gaps in diversity and in stereotyping theory regarding the presence and role of stereotyping in diverse teams, we examine how competence attributions of team members that stem from stereotypes shape an individual target team member’s behavior and individual performance in these teams. We argue that a target member’s task-gender congruence influences their behavioral dominance and individual performance via other team members’ competence attributions of the target team member, making our theory a cross-level theory. In addition, we argue that the relationship between task-gender congruence and competence attributions is moderated by the target member’s numeric representation (i.e., the extent to which the target team member is the solo representative of their gender in the team). Furthermore, we argue that the relationships between attributed competence and behavioral dominance and individual performance are moderated by the team’s diversity beliefs. The full research model is shown in Figure 1. In the following, we develop our arguments and hypotheses.

**Stereotypes and Competence Attributions in Gender-Diverse Teams**

Beyond the domains of diversity and stereotyping, there are a handful of studies that examined how stereotypes shape behavior and individual performance in diverse teams. These studies are grounded in research on expertise recognition (cf. Barton & Bunderson, 2014) and suggest that stereotypes inform members about who holds what expertise, and subsequently stimulate members to act upon those expectations. For example, Thomas-Hunt and Phillips (2004) investigated how a target member’s gender and
expertise affect other team members’ attributions of expertise and influence, and shape team-level performance on a masculine-typed task. The authors proposed and found that women are less influential than men with the same level of expertise, and that team performance suffers because women’s expertise was not used to the full in the team (for related studies, see Bunderson, 2003; Joshi, 2014; van Dijk et al., 2018).

These studies thus suggest that the content of stereotypes influences how members perceive and relate to each other (see Cuddy et al., 2007; Fiske et al., 2002). However, in focusing on team-level processes and team outcomes, these studies leave open the question how stereotypes shape interpersonal dynamics and individual members’ performance within teams. An exception is the study of Chatman et al. (2008), who examined other team members’ behavior and a target member’s individual performance in an experiment where teams of four members worked on masculine- or feminine-typed tasks. Chatman and colleagues found that task-gender congruence and a target member’s solo numeric representation (i.e., when the other three team members are of the opposite gender and the salience of the gender of the solo representative thus is likely to be high; Fiske & Taylor, 2008; Kanter, 1977; Sekaquaptewa & Thompson, 2003) predicted individual performance while controlling for individual ability. This effect was partially mediated by deferrals from other team members, thereby indicating that task-gender congruence influenced other members’ behavior toward solo representatives of their gender, which in turn affected the target member’s individual performance.

Importantly, the findings by Chatman et al. (2008) suggest that even in a team context, team members can hold more favorable attributions of target members of another social group than toward members of their own social group. As such, these findings contrast assumptions from conventional theory on diversity in teams about members of their own social group. Instead, they are in line with role congruity theory (Eagly & Karau, 2002) and the proportion of assertions versus hesitations and questions (Homan et al., 2020; van Knippenberg et al., 2004). Instead, they are in line with role congruity theory (Eagly & Karau, 2002) and the lack-of-fit model (Heilman, 1983) in suggesting that (stereotype-based) attributions of competence guide behavior and individual performance in (diverse) teams. Moreover, they are consistent with the argument of Fiske et al. (2002) and Abele et al. (2021) that people attribute competence on the basis of stereotype content (and stereotype fit in task settings) rather than on ingroup or outgroup membership. We therefore hypothesize that in team contexts, members attribute competence on the basis of a target member’s task-gender congruence, such that men as well as women team members give higher competence attributions to women (men) team members on feminine (masculine) tasks:

Hypothesis 1a: In gender-diverse teams, task-gender congruent team members receive higher competence attributions than task-gender incongruent team members.

Furthermore, we argue that numeric representation of a gender within a team will influence the extent to which task-gender congruence affects competence attributions. Since being a token representative of a gender increases category salience (Chatman et al., 2008; Heidemeier et al., 2021; Huguet & Regner, 2007; Sekaquaptewa & Thompson, 2003; cf. Kanter, 1977), we argue that the effect of task-gender congruence on competence attributions is strengthened when target team members are the solo representative of their gender in their team. We therefore hypothesize:

Hypothesis 1b: Numeric representation moderates the relationship between task-gender congruence and competence attributions, such that solo task-gender (in)congruent members receive higher (lower) competence attributions than nonsolo task-gender (in)congruent team members.

Stereotypes, Competence Attributions, Behavioral Dominance, and Individual Performance

Theory on the function of stereotypes suggests that competence attributions serve the function of knowing what to expect from target individuals and how to behave toward them (Abele et al., 2021; Fiske et al., 2002; Fiske & Neuberg, 1990; van Dijk et al., 2017). For example, the Behaviors from Intergroup Affect and Stereotypes (BIAS) map (Cuddy et al., 2007) suggests that competence attributions will result in attribution-consistent responses: Targets who are attributed high levels of competence are facilitated and given room to perform, whereas targets who are considered less competent are obstructed and worked against (cf. Chatman et al., 2008; Magee & Galinsky, 2008). In team contexts, such attribution-consistent responses are likely to be subtle and often para-verbal or nonverbal, such as the amount of interruptions, eye contact, management of interpersonal distance, and postural mimicry (Schermuly & Scholl, 2011).

Despite their subtlety, para-verbal and nonverbal interpersonal behaviors are likely to affect a target team member’s behavior (Meyer et al., 2016; Ridgeway et al., 1998). Research on the interpersonal circumplex (Costa et al., 2001; Wiggins, 1980) suggests that this is especially true for behavioral dominance, that is, the level of assertiveness and confidence in communication behavior. This is manifested through, for example, the frequency and length of speech, and the proportion of assertions versus hesitations and questions (cf. Loyd et al., 2010; Schermuly & Scholl, 2012; Schmid Mast, 2002). When individuals display dominance behaviors, others are more likely to display submissive behaviors, whereas when individuals display submissive behaviors, others are more likely to display dominance behaviors.

We argue that when other team members attribute high levels of competence to a target team member, they are likely to display submissive behaviors to facilitate a target and provide the target with room to perform (cf. Correll & Ridgeway, 2003; Cuddy et al., 2007; Wittenbaum & Bowman, 2005). For example, when there is a difficult question to be solved, team members are likely to gaze toward a target team member who they expect will know the answer (cf. Chatman et al., 2008), which stimulates the target team member to speak up and take the lead. We, therefore, hypothesize that when a task-gender congruent target team member receives higher
competence attributions from other members, the target team member will display more dominance behaviors:

**Hypothesis 2:** Other team members’ competence attributions mediate the positive relationship between a target team member’s task-gender congruence and the target team member’s dominance behavior.

Team members’ (stereotype-based) competence attributions of a target team member may also influence the target team member’s individual performance. Targets of stereotyping are often aware what others think of them (Ridgway et al., 1998), which tends to influence individual performance in a stereotype-confirming direction. Specifically, research on stereotype threat (Schmader et al., 2008; Steele & Aronson, 1995) and stereotype lift (Walston & Cohen, 2003) indicates that awareness of a task-relevant negative stereotype harms the individual performance of negatively stereotype typed targets, whereas it can lift the individual performance of nonstereotyped targets (e.g., men performing on a task that is congruent with the male stereotype). Moreover, research on stereotype susceptibility (Shih et al., 1999) shows that awareness of a task-relevant positive stereotype can enhance the individual performance of targets beyond their individual ability. Whereas such studies do not focus on team contexts, the mere presence of gender diversity in a gender-diverse team is likely to render task-gender stereotypes salient (Purdle-Vaughns et al., 2008; Turner et al., 1987), which in gendered task contexts thus is likely to result in higher competence attributions and subsequent individual performance for task-gender congruent targets.

In addition, other team members’ are likely to facilitate the performance of a task-gender congruent target by assigning a higher status to such targets. Research on status, which in team contexts often is determined by (attributed) competence (Correll & Ridgway, 2003), indicates that high-status team members tend to be the first to be given the room to take the lead and perform, offer suggestions, set directions, and their suggestions tend to be more positively evaluated (see also van Engen et al., 1996). Moreover, such target team members may engage less in social loafing and work harder because of their exemplary position (Chatman et al., 2008). In contrast, low-status team members tend to have more critical assessments of their suggestions and contributions, which renders them less likely to take the lead and more likely to show some reservations and follow the lead of others for reviews, see Correll & Ridgway, 2003; Magee & Galinsky, 2008; Wittenbaum & Bowman, 2005).

Based on these arguments, we hypothesize that in a gender-diverse team, other team members’ competence attributions enhance the performance of task-gender congruent targets:

**Hypothesis 3:** Other team members’ competence attributions mediate the positive relationship between a target team member’s task-gender congruence and the target team member’s individual performance.

Hypotheses 2 and 3 propose that members who are attributed more (less) competence by other team members behave more dominantly (submissively) and perform higher (lower). In combination, these hypotheses allude to a self-fulfilling tendency of stereotypes, such that competence attributions by team members result in corresponding behaviors and performance of the target individual that are likely to reinforce the initial competence attributions.

The danger of the self-fulfilling tendencies of gender-task stereotypes on dominance behavior and individual performance are that dominance behavior and individual performance in turn reinforce the same gender-task stereotypes, thereby institutionalizing gender stereotypes, sustaining the traditional gender segmentation of tasks and roles, and increasing the risk of inaccurate competence attributions (Jost & Banaji, 1994; van Dijk & van Engen, 2019). Furthermore, no single individual completely matches the stereotypes of that individual’s social group, which means that there is always some level of inaccuracy in stereotype-based attributions. In the following section, we propose that instilling pro-diversity beliefs among team members can block the propagation of stereotypes within work groups by stimulating group members to look beyond stereotype-based attributions of competence.

**Diversity Beliefs**

Because stereotypes constitute beliefs and assumptions regarding particular groups in a given context (Fiske & Taylor, 2008), a key resolution to eliminating the effects of stereotyping in teams may be to complement those beliefs with alternative beliefs that prevent team members from behaving on the basis of their stereotypes. We argue that this can be done by providing individuals with diversity beliefs, that is, “beliefs about the value of diversity to work group functioning” (van Knippenberg et al., 2007, p. 209).

Pro-diversity beliefs make team members aware that each individual may hold unique knowledge and perspectives that others are not aware of and provide meaningful contributions to the team effort. As such, pro-diversity beliefs carry the proposition that input from every team member, regardless of which member, can be of value. Team members with pro-diversity beliefs may still perceive a target team member as less competent because of stereotypes. However, because they also believe that contributions from such a target team member can be of value, they are less likely to act on stereotype-based competence attributions. We, therefore, argue that pro-diversity beliefs counteract the tendency to interpersonally behave in a stereotype-conforming way, such that team members will provide a target with room to perform regardless of the level of competence that is attributed to the target. Such room to perform will allow in particular targets who are attributed lower levels of competence to display more dominance behaviors and perform better.

In contrast, pro-similarity beliefs suggest that team members value the contributions of some members higher than of other members as they expect these members to provide better input.2

2 Whereas the preceding arguments suggest that behavioral dominance and individual performance are related, research on behavioral dominance also indicates that behavioral dominance and performance are distinct: Although behavioral dominance from a specific group member instills performance expectations toward this member among the other team members (Schmid Mast, 2002) and correlates with perceived expertise (Littlepage et al., 1995), the relationship between actual expertise and behavioral dominance is less clear: Whereas Chatman et al. (2008) findings suggest that team members who are perceived as competent by others—as implied through their deferrals—also perform better, Littlepage et al. (1995) found no significant correlation between dominance and actual expertise. Therefore, keeping diversity and performance separate allows us to investigate whether stereotypic gender-task congruence affects behavior, individual performance, or both in the team setting, and how these are related.
While the original theorizing surrounding these beliefs proposes that people attribute more value to similar individuals (Homan et al., 2007; van Knippenberg et al., 2007), we propose that at its core, the source of the belief in the higher or equal value of certain members, that is, being similar; that is, being a member of the same (sub-) group, does not matter. In suggesting that the input of some members is more valuable than those of others, pro-similarity beliefs convey the message that team functioning depends on a strongest-link principle, and that it therefore makes sense to determine (interpersonal) behavior on the basis of competence attributions. The assumed value of input from team members thus is more likely to be determined by their (stereotype-based) attributed task competence rather than similarity (Correll & Ridgeway, 2003; Magee & Galinsky, 2008). We, therefore, argue that in favoring the inputs from some team members over others, pro-similarity beliefs sustain or even reinforce the general tendency of team members to determine their (interpersonal) behavior on the basis of (stereotype-based) competence attributions, which inhibits in particular targets who are attributed lower levels of competence to display dominance behaviors and perform.

Taken together, we hypothesize that pro-diversity beliefs stimulate team members to be open to the possibility that (supposedly) less competent team members provide valuable input, causing members in diverse teams not to rely on stereotype-based competence attributions in directing their behavior toward target team members. In this way, pro-diversity beliefs weaken the stereotypic fit—attributions—behavior/performance link. In contrast, pro-similarity beliefs sustain or even reinforce the generic assumption that (supposedly) more competent team members provide more valuable input, causing members to determine their (interpersonal) behavior based on their competence attributions.

**Hypothesis 4a:** Diversity beliefs moderate the indirect effect of task-gender congruence on behavioral dominance via others’ competence attributions, such that the relationship between competence attributions and behavioral dominance is stronger under pro-similarity beliefs as compared to pro-diversity beliefs.

**Hypothesis 4b:** Diversity beliefs moderate the indirect effect of task-gender congruence on individual performance via others’ competence attributions, such that the relationship between competence attributions and individual performance is stronger under pro-similarity beliefs as compared to pro-diversity beliefs.

Finally, given that we hypothesized that being a solo representative of one’s gender strengthens the relationship between task-gender congruence and competence attributions, we assume that the relationships that we propose in Hypotheses 4a and b are stronger for solo individuals:

**Hypothesis 5a:** Numeric representation moderates the interaction effect of task-gender congruence and diversity beliefs on behavioral dominance, such that the relationship is stronger for solo task-congruent members than for nonsolo task-congruent members.

**Hypothesis 5b:** Numeric representation moderates the interaction effect of task-gender congruence and diversity beliefs on individual performance, such that the relationship is stronger for solo task-congruent members than for nonsolo task-congruent members.

In sum, our hypotheses constitute a cross-level model proposing that team features such as the task and the gender composition interact with team members’ individual attributes (i.e., their gender) and stereotypes, resulting in interpersonal competence attributions, which guide individual behavior.

**Method**

**Sample**

Three hundred and eighty-eight participants (186 men, 202 women, \(M_{age} = 26.88\) years, \(SD = 7.94\)) were randomly assigned to 97 teams of four in a 2 (congruence/incongruence between participant gender and task-gender stereotype: “typical” vs. “atypical”) \(\times\) 3 (group members’ numeric representation: solo, balanced, majority) \(\times\) 2 (pro-diversity beliefs vs. pro-similarity beliefs) design. Participants were recruited via the subject pool at the local (Swiss) university and received payment for participation as we explain below in further detail. Participants were mostly students from different fields of study and from different universities. In terms of ethnicity, all participants were white/Caucasian.

There were 28 teams with one woman and three men, 33 teams with two women and two men, and 36 teams with one man and three women. Forty-nine teams were assigned to the pro-diversity condition and 48 to the pro-similarity condition. Forty-nine teams worked on Emotional Intelligence (EI) problems, and 48 worked on math problems. Women who worked on EI problems and men who worked on math problems were categorized as having a gender that is congruent with the task stereotype (“typical”), the other participants were categorized as having a gender that is incongruent with the task stereotype (“atypical”). We reviewed math problems from the Graduate Management Aptitude Test (GMAT, e.g., Hecht & Schraeder, 1986) and selected corresponding German university entrance exam (“Abitur”) problems as stereotypically masculine problems, pertaining to the stereotype that men are usually perceived to excel in math (e.g., Huguet & Regner, 2007; Inzlicht & Ben-Zeev, 2000). For stereotypically feminine problems, we employed items from tests of EI because there is a stereotype pertaining to the higher emotional competence of women in comparison with men (Plant et al., 2000). German EI items were created based on items from self-scoring EI tests (e.g., Daniel, 2000) and the Mayer-Salovey-Caruso Emotional Intelligence Test; MSCEIT (Mayer et al., 2003.) One person from the team of authors who is a native German speaker translated the items, which were subsequently translated back into English by a native American academic proofreader who is fluent in German. Discrepancies were discussed and resolved in a group of German-speaking researchers and research assistants at the first author’s institution. Piloting these items as described in Appendix S1 revealed no issues. We established the validity of the EI and math task-gender stereotypes by conducting a pretest with a different sample; details are available from the online Supplemental S1 on the OSF repository for this manuscript at https://osf.io/3vw89. The pretest showed that individuals indeed believe that women score better on EI tasks than men, and that men score better on math tasks than women, thereby confirming the respective gender-task stereotypes.

Because all teams were gender diverse, teams always contained members whose gender stereotypically matched the task and...
members whose gender did not. Individual participants’ numeric representation was recorded as solo if he or she was the sole representative of the respective gender in the group. The other three members of the opposite gender were recorded as belonging to the majority. If there were two representatives of each gender in a team, their numeric representation was recorded as balanced. Diversity beliefs, as outlined below in further detail, were manipulated by informing participants about the supposed benefits of either team homogeneity or team heterogeneity.

Procedure

A random subsample of the students that were registered at the subject pool database of their university received an email with an invitation to participate in “a study on working together in teams.” Potential participants were offered a choice of experimental study credit or the equivalent of 25 USD for participation. Participants received an email invitation with time and date and a link to a prestudy online questionnaire that asked for gender and age and contained instructions for forming a code that participants also had to bring to their lab session for anonymous identification.

Upon arrival at the lab, participants had 10 min to individually work on paper-and-pencil tasks of the same type as in the upcoming team task. The leaflets containing these items contained more items than any participant was able to solve within the time limit. Speed and accuracy were used as indicators of individual ability (see below). Subsequently, participants were seated together at a round table and were asked to estimate each other’s task competence. For making these competence judgments, participants had to rely on overall first impressions of their fellow team members. Afterward, the team worked together on a single booklet containing further items of the same task type for 30 min and the interaction was recorded on video with three cameras pointed at the round table from different angles. Team members were instructed to work together as a team and discuss the potential answers for all items.

Prior to the team task but after the competence attributions, participants received a leaflet with information about the purpose of the study. This material was employed to manipulate participants’ diversity beliefs (see Homan et al., 2007, for a similar manipulation). In the pro-diversity condition, the text informed participants that prior research had shown that gender-diverse teams tend to perform better on the task at hand and that members of gender-diverse teams find working with each other on such tasks more enjoyable than members of gender-homogeneous teams. In the pro-similarity condition, participants were told the same with reference to gender-homogeneous teams. The texts for both conditions referenced teams of either team homogeneity or team heterogeneity.

At the time of the study, the Department of Psychology at the University of Zurich (where the study was conducted) employed a two-stage Internal review board (IRB) review process. In the first stage, authors had to self-assess their study according to an 11-item checklist querying potential ethical risks. Since we answered all questions with “no,” no formal IRB review, which would have comprised the second stage, was required. The checklist is available on the OSF repository.

Measures

Individual Ability

Prior to working on the team task, participants individually worked 10 min on a booklet containing 22 items that were of the same type as the items in the upcoming group task. Math items were scored according to the respective instructions and EI items were scored with the consensus method. Participants’ summed scores were subsequently Z-transformed.

Attributed Competence

Participants rated the competence in the given task at hand of each of their teammates from 0% (low competence) to 100% (high competence). Participants did not see the rating that they received. For each participant, we averaged the three ratings they received from the other team members into one score of attributed competence. This aggregation was justified as indicated by the two-way ICC(1) of .40, $F(380, 760) = 3.03, p < .001$, and by the two-way ICC(3) of .67.

Despite the high agreement among raters that justified aggregating these assessments to the individual level, we investigated whether rater gender and/or target gender also affected these ratings. For doing so, we constructed a multilevel data set with the three ratings that each participant gave out nested in participants, ICC(1) = 0.37, $F(375, 767) = 2.922, p < .001$. We regressed these given ratings on the Interaction of task (EI vs. math) × Target gender (male vs. female) × Rater gender (male vs. female) in an analysis of variance (ANOVA)-type mixed model, from which we extracted the estimated marginal means and plotted them, see Figure 2. The model revealed a significant main effect for target gender, $b = 8.50, 95\% CI [5.05; 11.93]$ and for the interaction of target gender and task type, $b = −17.14, 95\% CI [−22.12; −12.16]$, but no significant main effect or interactions for rater gender. As visible in Figure 2, both men and women raters rated female targets as more competent in EI tasks than male targets, and male targets as more competent in math tasks than female targets. However, “typical” women targets were evaluated more positively than “atypical” men, and “atypical” women targets more negatively than “atypical” men. This finding underscores the necessity to control for target gender in the analyses and we revisit this issue in the Discussion section.

Behavioral Dominance

Team members’ behavioral dominance was coded with the Discussion Coding System (DCS; Schermuly & Scholl, 2011, 2012) by trained coders. All videos were coded. For determining interrater reliability, five videos comprising 714 speech acts were chosen randomly and coded twice by two expert coders with prior experience with the DCS. Interrater reliability turned out to be
acceptable, ICC(1) = .65. After establishing acceptable interrater reliability, the remaining videos were coded by a single coder. For each participant, the dominance ratings associated with all of their speech acts were averaged into one measure of behavioral dominance that ranged from 2.00 to 3.66, $M = 2.93, SD = 0.27$.

**Individual Performance in the Team**

Individual performance was assessed through observational coding in a similar way as done by Chatman et al. (2008): Two independent coders (different from the ones doing the DCS coding) scored the answers that participants suggested and the answer the team reported on their answering form. Given that the math task and the EI task differed to the regard that math problems have a demonstrably correct solution whereas the EI items do not, scoring differed between tasks. For the math task, regardless of the number of times a participant mentioned an answer, each participant received a maximum of one point per response type per problem completed correctly by the team. Coders were instructed “to count an answer only if the participants actually stated an answer choice (A, B, C, D, or E) of their own accord” (Chatman et al., 2008, p. 149). In other words, a person received a point if they found or wrote down the correct answer individually and stated this choice verbally in their team (even if someone else had already given the same or a different answer) and if the team chose this answer as their team answer (independently of whether they chose it because of this particular participant or on the basis of what other participants said). Resulting individual math task scores ranged from 0 to 13 points.

For the EI task, coding was similar. Coders recorded the suggested answer option and the participant subsequently received the corresponding consensus EI score for that particular answer if the team settled on it. This consensus score is equivalent to the percentage of individuals in the referent sample who gave this answer; we provide more detail on this consensus score in Appendix S1 in the OSF repository. Resulting individual EI performance ranged from 1.29 to 8.94 points. Coders had a high agreement (Cohen’s $k$ for the math task = 0.96, Cohen’s $k$ for the EI task = 0.87).

In order to compare individual performances across tasks, we $Z$-transformed individual performance scores for each task separately before combining them into one individual-level performance variable, resulting in a variable where the 0 represents average individual performance for the given task.

**Manipulation Checks**

We checked the manipulation of diversity beliefs with four items in the posttask questionnaire: “Teams with members who are different from each other achieve higher performance than teams with members who are similar to each other,” “Teams with members who are different from each other experience more pleasant cooperation than gender-homogeneous teams,” “Teams with members from different genders achieve higher performance than gender-homogeneous teams,” and “Teams with members from different genders experience more pleasant cooperation than gender-homogeneous teams.” These items were presented with scales ranging from 1 (I strongly disagree) to 7 (I strongly agree).

**Results**

A subset of the data set has been used in another study (van Dijk et al., 2018), of which the only overlap in terms of main variables is the manipulation of diversity beliefs. The data transparency table, data, and analyses are available at the OSF repository accompanying this manuscript at https://osf.io/3wth9. Individual-level descriptives and bivariate correlations are given in Table 1.

**Level Issues and Analytical Strategy**

To account for the hierarchical structure of the data, we employed mixed-effects modeling (i.e., multilevel analyses) for testing the hypotheses (Bliese, 2000; Gelman & Hill, 2006). The data warranted this approach as task type and diversity beliefs were manipulated on the team level and because team membership explained significant levels of variance of dependent and mediating variables: While measures of individual performance were not nested in teams, ICC(1) < .05, $F(96, 291) < 1$, measures of attributed competence were, ICC(1) = .24, $F(96, 291) = 2.26, p < .001$, ICC(2) = .56, as was behavioral dominance, ICC(1) = .24, $F(96, 291) = 2.28, p < .001$, ICC(2) = .56.
Table 1
Means, Standard Deviations, and Correlations of Manipulations and Measurement Variables With Confidence Intervals (N = 388, Except for Age, n = 387)

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gender</td>
<td>0.48</td>
<td>0.50</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(female = 0, male = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Age (in years)</td>
<td>26.88</td>
<td>7.94</td>
<td>.18** [.09, .28]</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Individual ability (Z-transformed)</td>
<td>0.00</td>
<td>1.00</td>
<td>−.01 [−.11, .09]</td>
<td>−.15** [−.25, −.05]</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Task type (Math = −1, EI = 1)</td>
<td>0.01</td>
<td>1.00</td>
<td>.12* [.02, .22]</td>
<td>−.00 [−.10, .10]</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Attributed competence</td>
<td>59.84</td>
<td>11.88</td>
<td>.05 [−.05, .15]</td>
<td>.16** [.06, .26]</td>
<td>.21** [.11, .30]</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Numeric representation (solo = −1, balanced = 0, majority = 1)</td>
<td>0.33</td>
<td>0.74</td>
<td>−.08 [−.18, .02]</td>
<td>−.00 [−.10, .10]</td>
<td>−.10 [−.20, .00]</td>
<td>−.02 [−.12, .08]</td>
<td>−.07 [−.17, .03]</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Gender-task congruence (incongruent = −1, congruent = 1)</td>
<td>−0.01</td>
<td>1.00</td>
<td>−.02 [−.12, .08]</td>
<td>.06 [−.04, .16]</td>
<td>.20** [.10, .29]</td>
<td>.04 [−.06, .14]</td>
<td>.41** [.32, .49]</td>
<td>−.02 [−.12, .08]</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Diversity beliefs (DB, pro-similarity = −1, pro-diversity = 1)</td>
<td>0.01</td>
<td>1.00</td>
<td>−.11 [−.01, .09]</td>
<td>−.05 [−.14, .05]</td>
<td>.00 [−.10, .10]</td>
<td>−.03 [−.13, .07]</td>
<td>−.01 [−.11, .09]</td>
<td>.01 [−.09, .11]</td>
<td>.00 [−.10, .10]</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Attributed competence (Z-Transformed) × DB</td>
<td>−0.01</td>
<td>1.00</td>
<td>−.02 [−.12, .08]</td>
<td>.01 [−.09, .11]</td>
<td>.10 [−.00, .20]</td>
<td>−.12* [−.22, −.02]</td>
<td>.00 [−.10, .10]</td>
<td>.07 [−.03, .17]</td>
<td>.03 [−.07, .13]</td>
<td>.00 [−.10, .10]</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>10. Behavioral dominance</td>
<td>2.88</td>
<td>0.30</td>
<td>.26** [.17, .35]</td>
<td>.22** [.12, .31]</td>
<td>.03 [−.07, .13]</td>
<td>.18** [.08, .28]</td>
<td>.09 [−.01, .18]</td>
<td>−.01 [−.11, .09]</td>
<td>−.08 [−.18, .02]</td>
<td>−.01 [−.11, .09]</td>
<td>−.14** [−.23, −.04]</td>
<td>—</td>
</tr>
<tr>
<td>11. Individual performance (Z-transformed)</td>
<td>−0.00</td>
<td>1.00</td>
<td>.14** [.05, .24]</td>
<td>.00 [−.10, .10]</td>
<td>.30** [.20, .38]</td>
<td>.00 [−.10, .10]</td>
<td>.18** [.08, .27]</td>
<td>−.16** [−.26, −.07]</td>
<td>.16** [.06, .25]</td>
<td>.08 [−.02, .17]</td>
<td>.06 [−.04, .16]</td>
<td>.24** [.14, .33]</td>
</tr>
</tbody>
</table>

Note. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). EI = emotional intelligence.

* p < .05. ** p < .01.
We identified the most parsimonious structure of the random effects according to the recommendations by Bliwise (2016): We specified separate random-intercept models for the two focal dependent variables, individual performance and behavioral dominance, and one for the mediator attributed competence. For each of these three models, we used the focal individual-level predictor task-gender congruence as independent variable and specified each model twice: As a random-intercept model with a fixed slope for task typicality, and again as a random-intercept and slope model with a random slope for task typicality and compared model fits. In all three cases, the random-intercept and slope model did not fit to the data better than the random-intercept model: For individual performance, $\Delta \chi^2(3) = 1.271, p = .53$, for attributed competence, $\Delta \chi^2(3) = 3.683, p = .16$, and for behavioral dominance, $\Delta \chi^2(3) = 5.108, p = .08$. This implies that the amount of between-group variance of the relationship between task-gender congruence and these dependent variables was not large enough to warrant its inclusion in the models (Bliwise, 2016; Galecki & Burzykowski, 2013). Potentially, using a manipulation can lead to a rather homogenous structure of individual-level slope variation, which would explain the nonsignificant slopes between task typicality and the mediating and dependent variables. This does, however, not preclude the possibility of a moderation: The strength of this relationship may still be related to another variable (Bliwise, 2016). We thus used random-intercept models with fixed slopes for testing the hypotheses.

**Manipulation Checks**

In the pro-diversity condition, participants evaluated general team diversity as better for team performance, $M = 5.34, SD = 1.24$, and as more pleasant, $M = 4.77, SD = 1.39$, than in the pro-similarity condition, $M = 3.55, SD = 1.62$, and $M = 2.91, SD = 1.26$, respectively, $t(386) = 12.23, p < .001, d = 1.24$, and $t(386) = 13.84, p < .001, d = 1.41$. Likewise, participants in the pro-diversity condition evaluated gender diversity as better for team performance, $M = 5.62, SD = 1.24$, and as more pleasant, $M = 5.35, SD = 1.38$, than in the pro-similarity condition, $M = 3.52, SD = 1.66$, and $M = 3.67$, $SD = 1.59, t(386) = 14.10, p < .001, d = 1.43$, and $t(386) = 11.10, p < .001, d = 1.13$. We thus deemed the manipulation a success.

**Task-Gender Congruence and Attributed Competence**

Hypothesis 1a predicted that stereotypical task-gender congruence results in higher levels of competence being attributed to the target by the other team members. To test it, we constructed a random-intercept mixed model regressing received competence attributions on dummy-coded stereotypical task-gender congruence ($1 = $ typical, i.e., the target’s gender fitted to the task, $0 = $ atypical otherwise), controlling for target’s individual task ability, age, and gender, which we allowed to interact with task-gender congruence. The model revealed a significant main effect of gender such that women received 3.51 points less than men, 95% CI $[-9.13; -2.82]$, and—in support of Hypothesis 1a—a main effect of task typicality such that team members whose gender matched the task received 4.85 points more than those team members whose gender did not, 95% CI $[1.38; 8.34]$. The model also revealed a significant interaction between target gender and task typicality, $b = 9.18, 95\%$ CI $[3.48; 14.87]$. To interpret it, we extracted and plotted the model’s estimated marginal means, see Figure 3. It shows that the difference between “atypical” and “typical” women is larger (14.03 points) than for “atypical” and “typical” men (4.85 points). Importantly, planned contrasts show that within each gender, the difference between “atypical” and “typical” targets was significant, both $p < .01$. The model accounted for 21% of the variance of competence attributions. The standardized effect size of the difference in competence attributions between atypical and typical men—which is interpretable analogous Cohen’s $d$—was 0.45 (i.e., a small effect close to medium). For women, the difference was equivalent to an effect of 1.29; that is, a large effect.

Hypothesis 1b predicted that this relationship is moderated by participants’ numeric representation in their team. To test it, we added a dummy-coded variable to the previous model that was coded as 1 for solo participants who were the only representatives of their gender in their team and as 0 for all other nonsolo participants. We added this variable as a main effect and as interacting with participants’ stereotypical task-gender congruence. The confidence interval of the main effect of numeric representation on attributed competence included 0, as did all interactions containing this predictor. We thus rejected Hypothesis 1b.

**Multilevel Path Model**

Taken together, Hypotheses 2 and 3 predict that others’ competence attributions mediate between target’s stereotypical task-gender congruence on the one side and their behavioral dominance (Hypothesis 2) and individual performance (Hypothesis 3) on the other. Hypotheses 4a and 4b predict that these two indirect paths are moderated by diversity beliefs, and Hypotheses 5a and 5b add participants’ solo status as a further moderator to the indirect effect (see also Figure 1). To test all of these proposed relationships in a parsimonious way while controlling for all other relationships, we conducted a multilevel moderated mediation analysis (Preacher et al., 2010) by specifying an according to multilevel path model with the lavaan R package (Rosseel, 2012), Version 0.6-7, following the logic provided by Preacher et al. (2007). It contained behavioral dominance and individual performance as correlated outcomes as—in the terminology of Stapleton et al. (2016)—observed variables
with a within- and a between-team part. In other words, the model decomposed the variance of these variables into a within-team part and a between-team part, which is equivalent to treating them as having random intercepts (Stapleton et al., 2016). We entered stereotypical task-gender congruence as an individual-level effect-coded predictor of the individual-level components of attributed competence, behavioral dominance, and individual performance, coded as −1 for “atypical” and 1 for “typical.” We added the control variables individual task ability and age as variables with a within-and-between component. Participant gender was entered as an individual-level only variable while controlling for team gender composition on the team level by adding the number of men team members to the between-team level. Given that our manipulation of diversity beliefs occurred on the team level, we also constrained it on the between-team level in the model, coded as −1 for pro-similarity beliefs and 1 for pro-diversity beliefs. The variable representing the interaction between attributed competence and diversity beliefs was entered with a within- and a between-component. All parameter estimates of the model, including the indirect moderated effects, are given in Table 2 and the model is summarized in Figure 4.3.

Estimates of the model, including the indirect moderated effects, are entered with a within- and a between-component. All parameter estimates of the model, including the indirect moderated effects, are given in Table 2 and the model is summarized in Figure 4.3. Relationships on the individual level thus represent fixed effects and team-level effects need to be interpreted as relationships between team-level variance components (i.e., as covariances among random intercepts). For example, $\beta_{fi} b_{11w}$ denotes the between-team relationship between the between-team variance of the number of men team members and the between-team variance of attributed competence.

Before we test the remaining hypotheses on the basis of the path model, we investigated its overall fit on the basis of the criteria by Schermelleh-Engel et al. (2003). The model exhibited excellent fit, $\chi^2 (df = 12) = 7.068, p = .853, \text{RMSEA} = 0.000, 95\% \text{CI} [0.000, 0.029], \text{NFI} = 0.972, \text{CFI} = 1, \text{TLI} = 1, \text{SRMR}_{\text{a}} = 0.020, \text{SRMR}_{\text{b}} = 0.065$.

Hypothesis 2 predicted that competence attributions mediate the relationship between stereotypical task-gender congruence and behavioral dominance while controlling for individual ability. Contrary to the hypothesis, the unstandardized indirect effect was not significant, $a_w \times b_{11w} = 0.018, 95\% \text{CI} [-0.033, 0.070]$. The model also did not reveal a significant direct relationship between individual ability and behavioral dominance. It only showed a relationship between gender and behavioral dominance such that men behaved more dominantly than women, $\beta = .25, p < .01$. We thus rejected Hypothesis 2.

Hypothesis 3 stated that competence attributions mediate the relationship between stereotypic gender-task-fit and individual performance while controlling for individual ability. In support of the hypothesis, the corresponding indirect effect was significant, $a_w \times b_{21w} = 0.062, 95\% \text{CI} [0.007, 0.116]$.

Hypothesis 4a proposed that diversity beliefs moderate the relationship between stereotypical task-gender congruence, competence attributions, and behavioral dominance, while Hypothesis 4b did the same with individual performance as dependent variable while controlling for individual ability. We tested these hypotheses by estimating the proposed indirect effect under the different levels of the moderator with the formulas specified by Preacher et al. (2007). To test whether diversity beliefs moderate the indirect effect of stereotypical task-gender congruence on dominance via competence attributions (Hypothesis 4a), we thus investigated the mediation under the two levels of the moderator. For pro-diversity beliefs, this indirect effect is equivalent to $a_w \times (b_{11b} + (1 \times b_{22b}))$ (Preacher et al., 2007), which resulted in an estimate of $-0.163, 95\% \text{CI} [-0.349, 0.024]$. For pro-similarity beliefs, this path was significant, $a_w \times (b_{11b} + (-1 \times b_{22a})) = 0.178, 95\% \text{CI} [0.004, 0.352], \beta = .235$. Hypothesis 4a was thus supported by the data.

We found a similar pattern for the moderation of the indirect path from stereotype gender-task fit via attributed competence on individual performance: Under pro-diversity beliefs, the indirect effect was not significant, $a_w \times (b_{11w} + 1 \times b_{22w}) = 0.000, 95\% \text{CI} [-0.116; 0.116]$. Under pro-similarity beliefs, the confidence interval of the indirect path excluded 0, $a_w \times (b_{11w} - 1 \times b_{22a}) = 0.123, 95\% \text{CI} [0.003; 0.243], \beta = .320$. The data thus supported Hypothesis 4b. In the absence of the initial moderation of numeric representation between task typicality and competence attributions, we rejected Hypotheses 5a and 5b. As visible in the standardized residual variance components in Table 2, the model explained 14% of the variance of behavioral dominance and 11% of the variance of individual performance.

**Discussion**

The results confirmed our expectation that stereotypes in gender-diverse teams tend to yield stereotype-confirming behaviors and individual performance. Overall, when team members’ belief in the value of gender diversity is not taken into account, we found that individual group members whose gender was congruent with the task stereotype (e.g., men—math, women—emotion recognition) were attributed higher competence and outperformed other group members above and beyond their established individual ability in math and emotion recognition. Moreover, when teams did not believe in the value of diversity, other team members’ competence attributions of a target member also translated into dominance such that supposedly competent target members behaved more dominantly. These results were independent of team gender composition. In line with our arguments and expectations, we found that pro-diversity beliefs mitigate these stereotyping-confirming behaviors and individual performance tendencies in gender-diverse teams.

**Stereotyping in Gender-Diverse Work Groups**

Our findings strongly oppose the predictions from research on team diversity and relational demography from the social categorization perspective, which suggest that perceiving another team member as different members in negative attributions and performance consequences (e.g., Carton & Cummings, 2012; Homan et al., 2020; van Knippenberg et al., 2004). We found that team members whose gender differs from the other team members can profit from being different if their gender stereotypically matches the task at hand. In such circumstances, their difference boosts their individual performance beyond what one would expect in light of their individual task ability. This stereotype lift is not only mirrored in “typical” team members’ individual performance, but—under pro-similarity beliefs—also in their behavioral dominance. In both cases, we provide a rare and important observation of a stereotype lift effect in a team context, and found that an interpersonal process drove these effects: Target team members’ behavior and individual diversity beliefs.

---

3 Note that we brought other team members’ competence attributions to the individual level by aggregating all received ratings to the respective team member, which is why the path model only contains two levels, individual and team.
### Table 2

*Multilevel Path Model Parameter Estimates (N = 387)*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Label</th>
<th>(b)</th>
<th>CI LL</th>
<th>CI UL</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual level (focal team member)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Direct effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task ability</td>
<td>(\rightarrow) Attributed competence</td>
<td>0.013</td>
<td>−0.083</td>
<td>0.109</td>
<td>0.015</td>
</tr>
<tr>
<td>Gender (0 = female, 1 = male)</td>
<td>(\rightarrow) Attributed competence</td>
<td>0.093</td>
<td>−0.075</td>
<td>0.261</td>
<td>0.054</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>(\rightarrow) Attributed competence</td>
<td>−0.010</td>
<td>−0.022</td>
<td>0.002</td>
<td>−0.087</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>(\rightarrow) Attributed competence</td>
<td>(a_{w})</td>
<td>0.406</td>
<td>0.324</td>
<td>0.489</td>
</tr>
<tr>
<td>((-1 = incongruent, 1 = congruent))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task ability</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>0.076</td>
<td>−0.029</td>
<td>0.181</td>
<td>0.083</td>
</tr>
<tr>
<td>Gender (0 = female, 1 = male)</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>0.437</td>
<td>0.252</td>
<td>0.621</td>
<td>0.250</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>0.017</td>
<td>0.004</td>
<td>0.030</td>
<td>0.151</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>−0.099</td>
<td>−0.202</td>
<td>0.004</td>
<td>−0.114</td>
</tr>
<tr>
<td>((-1 = incongruent, 1 = congruent))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributed competence</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>(b_{13w})</td>
<td>0.045</td>
<td>−0.081</td>
<td>0.171</td>
</tr>
<tr>
<td>Attributed competence (\times) Diversity beliefs</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>(b_{12w})</td>
<td>−0.045</td>
<td>−0.155</td>
<td>0.065</td>
</tr>
<tr>
<td>Task ability</td>
<td>(\rightarrow) Individual performance</td>
<td>0.286</td>
<td>0.179</td>
<td>0.394</td>
<td>0.274</td>
</tr>
<tr>
<td>Gender (0 = female, 1 = male)</td>
<td>(\rightarrow) Individual performance</td>
<td>0.335</td>
<td>0.130</td>
<td>0.540</td>
<td>0.168</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>(\rightarrow) Individual performance</td>
<td>0.005</td>
<td>−0.009</td>
<td>0.018</td>
<td>0.037</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>(\rightarrow) Individual performance</td>
<td>0.043</td>
<td>−0.067</td>
<td>0.152</td>
<td>0.043</td>
</tr>
<tr>
<td>((-1 = incongruent, 1 = congruent))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributed competence</td>
<td>(\rightarrow) Individual performance</td>
<td>(b_{21w})</td>
<td>0.151</td>
<td>0.019</td>
<td>0.284</td>
</tr>
<tr>
<td>Attributed competence (\times) Diversity beliefs</td>
<td>(\rightarrow) Individual performance</td>
<td>(b_{22w})</td>
<td>0.061</td>
<td>−0.054</td>
<td>0.176</td>
</tr>
<tr>
<td><strong>Covariances among dependent and mediating variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual performance</td>
<td>(\leftrightarrow) Behavioral dominance</td>
<td>0.195</td>
<td>0.110</td>
<td>0.281</td>
<td>0.257</td>
</tr>
<tr>
<td>Attributed competence (\times) Diversity beliefs</td>
<td>(\leftrightarrow) Attributed competence</td>
<td>0.060</td>
<td>−0.017</td>
<td>0.137</td>
<td>0.091</td>
</tr>
<tr>
<td><strong>Residual variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributed competence</td>
<td>(\leftrightarrow) Attributed competence</td>
<td>0.571</td>
<td>0.478</td>
<td>0.665</td>
<td>0.770</td>
</tr>
<tr>
<td>Behavioral dominance</td>
<td>(\leftrightarrow) Behavioral dominance</td>
<td>0.680</td>
<td>0.569</td>
<td>0.790</td>
<td>0.894</td>
</tr>
<tr>
<td>Individual performance</td>
<td>(\leftrightarrow) Individual performance</td>
<td>0.850</td>
<td>0.728</td>
<td>0.971</td>
<td>0.857</td>
</tr>
<tr>
<td><strong>Covariances among independent variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributed competence (\times) Diversity beliefs</td>
<td>(\leftrightarrow) Attributed competence (\times) Diversity beliefs</td>
<td>0.760</td>
<td>0.636</td>
<td>0.883</td>
<td>1.000</td>
</tr>
<tr>
<td>Gender (0 = female, 1 = male)</td>
<td>(\leftrightarrow) Task ability</td>
<td>−0.016</td>
<td>−0.068</td>
<td>0.036</td>
<td>−0.034</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>(\leftrightarrow) Task ability</td>
<td>−1.512</td>
<td>−2.361</td>
<td>−0.663</td>
<td>−0.209</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>(\leftrightarrow) Task ability</td>
<td>(b_{192})</td>
<td>0.192</td>
<td>0.092</td>
<td>0.291</td>
</tr>
<tr>
<td>((-1 = incongruent, 1 = congruent))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (in years)</td>
<td>(\leftrightarrow) Gender (0 = female, 1 = male)</td>
<td>(0.662)</td>
<td>(0.242)</td>
<td>(1.082)</td>
<td>(0.174)</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>(\leftrightarrow) Gender (0 = female, 1 = male)</td>
<td>−0.004</td>
<td>−0.054</td>
<td>0.046</td>
<td>−0.008</td>
</tr>
<tr>
<td>((-1 = incongruent, 1 = congruent))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>(\leftrightarrow) Age (in years)</td>
<td>0.425</td>
<td>−0.349</td>
<td>1.199</td>
<td>0.056</td>
</tr>
<tr>
<td>((-1 = incongruent, 1 = congruent))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Team level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Direct effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task ability</td>
<td>(\rightarrow) Attributed competence</td>
<td>0.303</td>
<td>−1.043</td>
<td>1.648</td>
<td>0.182</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>(\rightarrow) Attributed competence</td>
<td>0.119</td>
<td>−0.134</td>
<td>0.272</td>
<td>0.468</td>
</tr>
<tr>
<td>Number of male team members</td>
<td>(\rightarrow) Attributed competence</td>
<td>(a_{b})</td>
<td>−0.036</td>
<td>−0.267</td>
<td>0.194</td>
</tr>
<tr>
<td>Task ability</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>−0.155</td>
<td>−1.649</td>
<td>1.339</td>
<td>−0.098</td>
</tr>
<tr>
<td>Number of male team members</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>−0.024</td>
<td>−0.274</td>
<td>0.227</td>
<td>−0.039</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>0.152</td>
<td>−0.206</td>
<td>0.510</td>
<td>0.630</td>
</tr>
<tr>
<td>Attributed competence</td>
<td>(\rightarrow) Behavioral dominance</td>
<td>(b_{11b})</td>
<td>−0.060</td>
<td>−0.907</td>
<td>0.787</td>
</tr>
</tbody>
</table>

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
<table>
<thead>
<tr>
<th>Effect</th>
<th>Label</th>
<th>b</th>
<th>CI LL</th>
<th>CI UL</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity beliefs (−1 = pro-similarity; 1 = pro-diversity)</td>
<td>→ Behavioral dominance</td>
<td>0.010</td>
<td>−0.109</td>
<td>0.130</td>
<td>0.021</td>
</tr>
<tr>
<td>Attributed competence × Diversity beliefs</td>
<td>→ Behavioral dominance</td>
<td>( b_{12b} )</td>
<td>−0.446</td>
<td>−0.887</td>
<td>−0.004</td>
</tr>
<tr>
<td>Task ability</td>
<td>→ Individual performance</td>
<td>0.236</td>
<td>−0.282</td>
<td>0.754</td>
<td>0.528</td>
</tr>
<tr>
<td>Number of male team members</td>
<td>→ Individual performance</td>
<td>−0.097</td>
<td>−0.238</td>
<td>0.044</td>
<td>−0.574</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>→ Individual performance</td>
<td>0.005</td>
<td>−0.087</td>
<td>0.998</td>
<td>0.079</td>
</tr>
<tr>
<td>Attributed competence</td>
<td>→ Individual performance</td>
<td>−0.030</td>
<td>−0.345</td>
<td>0.284</td>
<td>−0.113</td>
</tr>
<tr>
<td>Diversity beliefs (−1 = pro-similarity; 1 = pro-diversity)</td>
<td>→ Individual performance</td>
<td>0.075</td>
<td>−0.020</td>
<td>0.170</td>
<td>0.551</td>
</tr>
<tr>
<td>Attributed competence × Diversity beliefs</td>
<td>→ Individual performance</td>
<td>( b_{23b} )</td>
<td>−0.152</td>
<td>−0.407</td>
<td>0.103</td>
</tr>
</tbody>
</table>

Covariances among dependent and mediating variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual performance</td>
<td>Behavioral dominance</td>
</tr>
<tr>
<td>Diversity beliefs (−1 = pro-similarity; 1 = pro-diversity)</td>
<td>Attributed competence × Diversity beliefs</td>
</tr>
<tr>
<td>Attributed competence × Diversity beliefs</td>
<td>Attributed competence</td>
</tr>
</tbody>
</table>

Residual variances

<table>
<thead>
<tr>
<th>Variable</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual performance</td>
<td>Individual performance</td>
</tr>
<tr>
<td>Attributed competence</td>
<td>Attributed competence</td>
</tr>
<tr>
<td>Behavioral dominance</td>
<td>Behavioral dominance</td>
</tr>
</tbody>
</table>

Covariances among independent variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (in years)</td>
<td>Task ability</td>
</tr>
<tr>
<td>Number of male team members</td>
<td>Task ability</td>
</tr>
<tr>
<td>Number of male team members</td>
<td>Age (in years)</td>
</tr>
</tbody>
</table>

Indirect effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>Attributed competence</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>Attributed competence</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>Attributed competence</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>Attributed competence</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>Attributed competence</td>
</tr>
<tr>
<td>Stereotypic gender-task congruence</td>
<td>Attributed competence</td>
</tr>
</tbody>
</table>

Note. Unless otherwise specified in brackets, all measurement variables are Z-transformed. Bold values denote parameters whose 95% confidence interval excludes 0.
performance was affected by how competent the other team members perceived them to be. These findings also underscore the importance of taking a cross-level perspective when investigating processes in diverse teams.

From a functional perspective, this finding illustrates the potential utility of stereotypes (Fiske & Neuberg, 1990; van Dijk et al., 2017). Congruent team members on average had higher ability levels than noncongruent members, which makes relying on the competence of congruent team members in the absence of any other competence-identifying information a safer bet than relying on incongruent members. This functionality of stereotypes contrasts most research focusing on the negative consequences of stereotypes (Eagly & Sczesny, 2019; Heilman & Caleo, 2018; Koch et al., 2015; Rudman & Phelan, 2008), but is in line with research on stereotype accuracy (Jussim et al., 2015), and explains why people continue to rely on stereotypes despite a multitude of efforts to eradicate them from the workplace.

However, despite this potential functionality of stereotypes, our findings also reveal that stereotypes shape attributions of competence and corresponding dominance behaviors (under pro-similarity beliefs) and individual performance in line with the stereotype-based expectations, thereby indicating that stereotypes tend to be behaviorally confirmed in gender-diverse teams. We argue that this finding is worrisome for two reasons. The first is theoretical in nature because decades of research on diversity in work groups have tended to neglect theory on stereotyping (cf. Harrison & Klein, 2007; van Knippenberg et al., 2004), which made the processes and dynamics in team settings that stereotypes set in motion invisible. Instead, theory on diversity in work groups has had a myopic focus on ingroup favoritism and subgroup formation (cf. Carton & Cummings, 2012) as well as a strong focus on team-level processes and outcomes (Meyer, 2017). Furthermore, there has been ample theory on the negative consequences of stereotyping on the behavior and individual performance of individual employees (Cuddy et al., 2011; Schmader et al., 2008), but such research has not considered how the wider work group or team context affects the influence and consequences of stereotypes. Arguably, based on the contact hypothesis (cf. Allport, 1954), the literature on diversity in work groups—as well as the literature on stereotyping—assumes that in teams, members get to know each other better and will overcome making inferences and behaving based on stereotypes (cf. Bunderson, 2003; Harrison et al., 2002). Our findings challenge this assumption, and suggest that stereotypes instead can have a pervasive, self-reinforcing effect on attributions and behavior.
in diverse teams (cf. van Dijk et al., 2017), thereby calling for a reconsideration of some fundamental assumptions underlying research on diversity in teams.

Given the short-lived nature of our experiment, our study calls for more research on the long-term consequences of teamwork for reinforcing gender stereotypes. In particular, the assumption that ingroup–outgroup dynamics and ingroup favoritism shape attributions and behaviors in diverse teams requires a reassessment. If that were true, then it would not matter for attributions of competence and behavioral dominance if a gender-diverse team would work on a masculine or a feminine task. However, we find that those tasks yield opposing dynamics: Whereas on masculine tasks men are attributed higher levels of competence, behave more dominantly (under pro-similarity beliefs) and outperform women, we find the reverse on feminine tasks, such that women are attributed higher levels of competence, behave more dominantly (under pro-similarity beliefs) and outperform men. These dynamics are not a matter of a more dominant social group (e.g., men) favoring the ingroup over others (e.g., women), but instead of all members favoring those individuals whose gender stereotypically fits the task context better. Whereas men on masculine tasks and women on feminine tasks thus tend to display ingroup favoritism, on tasks that do not stereotypically fit their gender men and women display outgroup favoritism.

In contrast to the current assumption that social categorization is used as a means to differentiate between in- and outgroup members (cf. van Knippenberg et al., 2004), our findings indicate that team members use social categorization to make an assessment of other team members’ worth. When the combination of gender stereotypes and stereotypes regarding the task context suggest that another team member is competent (cf. Eagly & Karau, 2002), then team members will defer more to that team member (cf. Chatman et al., 2008) and behave more submissively. In contrast, if the combination of gender and task stereotypes suggest that a target member is less competent, then team members will behave more dominantly and be more critical toward input from the target member (cf. Wittenbaum & Bowman, 2005). Importantly, such dominance behaviors relate to elements that are central in information elaboration (e.g., information sharing, critically discussing shared information), which is crucial for the performance of diverse teams (Srikanth et al., 2016; van Knippenberg et al., 2004). We, therefore, argue that stereotype-related processes require a more central role in theory on team diversity.

A second reason why our conclusions are worrisome is more social in nature. Specifically, our findings suggest that instead of overcoming stereotype-based impressions, stereotypes in a team context tend to be reinforced and legitimized in gender-diverse work teams under pro-similarity beliefs. Because individuals’ behaviors represent the idiosyncratic signals that people use to look beyond stereotype-based perceptions and arrive at more accurate attributions of competence (Fiske & Neuberg, 1990), the behavioral confirmation of stereotypes can be easily perceived as a justification of the stereotypes. As such, there is a danger of stereotypes creating a vicious cycle, in which the behavioral confirmation of a stereotype in turn reinforces the stereotype, which results in a further behavioral confirmation and, over time, institutionalization, legitimation, and normalization of the stereotype (cf. Jost & Banaji, 1994; Rudman & Fairchild, 2004; van Dijk & van Engen, 2019).

The danger of competence attributions shaping individual behavior that confirms stereotypes in a team context is in particular illustrated by our finding that stereotypes shape individual performance above and beyond one’s individual ability in a stereotype-confirming way. Our findings show that gender stereotypes in diverse teams enable task-typical members to not only outperform their task-atypical counterparts, but also themselves. The gender-diverse composition of the teams in combination with the stereotypes thus do not just behaviorally confirm any potential actual differences between the genders, but enlarge those differences. The diverging individual performance differences between task-typical and task-atypical members also makes it easier to notice individual performance differences, thereby contributing to the likelihood that stereotypes become justified and institutionalized (van Dijk et al., 2020). As such, in contrast to the common assumption that the influence of stereotypes tends to wane in team contexts (e.g., Harrison et al., 2002), our findings suggest that team contexts make stereotypes thrive, and highlight the importance of studying the influence of stereotypes in team contexts.

Inhibiting the Self-Reinforcing Effects of Stereotypes by Pro-Diversity Beliefs

To overcome the tendency that stereotypes are behaviorally confirmed in gender-diverse teams, we have argued that instilling pro-diversity beliefs may be helpful by motivating people not to act based on their stereotypes. In support of this argument, our findings show that the consequences of stereotypes are malleable by instilling pro-diversity beliefs. We found this to be true for behavioral dominance as well as individual performance: Whereas under pro-similarity beliefs, congruent members behaved more dominantly and outperformed incongruent members, these effects disappeared under pro-diversity beliefs. Given that diversity beliefs were manipulated after participants made initial (stereotype-) based competence attributions, this finding indicates that pro-diversity beliefs inhibit the effects of stereotype-based competence attributions. The assertion that diversity beliefs can prevent the application of stereotypes is a stark deviation from their prior use in the diversity literature (e.g., Homans et al., 2015; van Dick et al., 2008). Instead of discussing group-level outcomes, we look at the interpersonal dynamics underlying the group-level effects of having pro-diversity or pro-similarity beliefs.

We believe this finding provides reason for optimism. It suggests that pro-diversity beliefs motivate team members to reconsider their competence assumptions and make them aware that team members who are perceived as less competent may still be valuable to the group. As such, pro-diversity beliefs pave the way for less stereotype-confirming behaviors and, hence, more accurate attributions and impression formation. Moreover, in a field that is predominated with findings on the pervasive, negative consequences of stereotypes, this is a refreshing insight that enables researchers as well as practitioners to consider how to overcome the negative consequences of stereotyping. Given that the pro-diversity beliefs manipulation in our study resembles the diversity statements that companies are increasingly making, it may also represent a manipulation that is relatively easy to use in practice. However, because the effectiveness of diversity statements also relies on a number of factors (Carnes et al., 2019), we first call for future research to further assess how pro-diversity beliefs can be effectively installed in the workplace.

In the absence of a control condition without a diversity beliefs manipulation, our findings remain mute about the default level of diversity beliefs. However, at least with regard to performance, the
findings by Chatman et al. (2008), whose study did not include diversity beliefs, resemble our findings under the pro-similarity beliefs condition. This suggests that the default team setting is likely one where team members’ dominance and performance is influenced by their tendency to act on the basis of stereotypes.

**Suggestions for Future Research**

In general, our study supports the suggestion that (stereotype-based) competence attributions are the mechanism between stereotyping on the one hand, and interpersonal and group dynamics and outcomes on the other hand (van Dijk et al., 2017). In showing the pervasive effects of stereotypes on behavior and individual performance, our study thus emphasizes the importance for future research on diversity in teams to examine how stereotypes shape attributions, behavior, dynamics, and outcomes. Given the scarcity of studies on stereotypes in (diverse) team contexts, there are ample opportunities here, and in particular studies that further examine the extent to which stereotyping relates to outgroup favoritism and thus challenges a fundamental assumption of current diversity theory can be highly relevant. Our study solely focused on gender diversity, but our theory and arguments are not necessarily bound to gender diversity. A logical step forward therefore would be for future studies to examine stereotype-related processes in teams that are diverse on dimensions other than gender.

A somewhat related implication for future diversity research pertains to taking into account the composition of the team. Whereas team members whose demographics are congruent with the task stereotype tend to profit from a team’s heterogeneity, other members may suffer from it. For example, our findings suggest that interpersonal dynamics tend to be vastly different in a team with three women and one man compared to a team with three men and one woman. Such contrasting effects within gender-diverse teams may not be visible if one only looks at team diversity and not at team composition, which is common practice in diversity research (Harrison & Klein, 2007; van Dijk et al., 2012; for an exception, see Bell, 2007). Therefore, accounting for the idea that team diversity may have differential effects on team members depending on their stereotype-task congruence within the configuration of their team with multilevel theory and methods may be one way of advancing research on team diversity further.

In that regard, it is interesting that our manipulation of numeric representation did not yield any effect. Whereas in prior studies differences between gender-diverse team members were only—or especially—visible for solo representatives of their gender (e.g., Chatman et al., 2008; Sekaquaptewa & Thompson, 2002, 2003), in our study differences were noticeable between men and women team members regardless of their numeric representation. In essence, we believe this nonsignificance of numeric representation speaks to the robustness of the effects that we found of stereotypes in gender-diverse teams.

A surprising finding was that “typical” women were attributed higher levels of competence than “typical” men, whereas “atypical” women were attributed lower levels of competence than “atypical” men. These findings differ somewhat from the symmetry that is predicted by role congruity theory (Eagly & Karau, 2002) and the lack-of-fit model (Heilman, 1983) regarding the consequences of typicality and atypicality being similar for women and men. Furthermore, these findings also differ from the burgeoning work on “atypical” men, which suggests that “atypical” men may be as susceptible for negative attributions as “atypical” women—if not more (Manzi, 2019; Moss-Racusin & Johnson, 2016). Potential explanations for these surprising findings may lie in the nature of the tasks employed in this specific study (e.g., math and EI being on the extreme ends of task-gender typicality), the presence of stronger competence stereotypes regarding women compared to men, or women overall being stereotyped more than men. However, any further argumentation for this finding would not go beyond mere speculation. We, therefore, call for future research to replicate and help account for this unexpected finding.

**Practical Implications**

Our findings emphasize that diverse team contexts tend to activate stereotypes and their behavioral consequences. Because such corresponding behaviors tend to confirm the initial stereotypes, our study suggests that gender-diverse teams may provide a context where gender stereotypes are justified and legitimized (van Dijk et al., 2020).

A first implication for practitioners thus yields a warrant against organizational contexts and settings that increase stereotype salience. However, we are aware that often this is just impossible, and that our findings in part point at the functionality of stereotypes also suggests that stereotypes are unlikely to disappear altogether. Rather than trying to eradicate stereotypes, we therefore argue that practitioners would do well to try to mitigate the detrimental effects of stereotyping. Our findings indicate that making members aware of the benefit of diversity and the utility of task-atypical members (e.g., Joshi, 2014) may be enough to establish pro-diversity beliefs in an organization and hence eradicate the negative behavioral consequences of stereotypes.

The fact that diversity beliefs can counter the behavioral effects of stereotypes thus shows that highlighting the value of diversity is a viable alternative to other forms of communication that aims to counter stereotypes. For example, the portrayal of successful non-typical role models such as female math professors has shown to impair women’s interests in nonstereotypical work domains (Betz & Sekaquaptewa, 2012; see also Dobbin & Kalev, 2016). Highlighting the value of gender diversity at the workplace thus may be more helpful in enhancing the proportion of atypical team members than portraying individual success stories, and in making them thrive.

**Strengths, Limitations, and Outlook**

In our study, we were able to observe actual behavior and individual performance of participants working together on a task in a controlled way, which allowed us to disentangle the effects of gender and stereotype-task fit in different diversity belief contexts while controlling for individual abilities. Furthermore, our experimental design helped us examine the causality chains of task context, attributions, and behavior and individual performance in diverse teams.

The strengths of this context come with some limitations. Specifically, it took place in an artificial setting with a specific participant group, working on a task that is disconnected from their everyday life, and collaborating with fellow participants for which there is no expectation of a continued collaboration or future work relation.
Conclusion

Our findings illustrate the prevalence and persistence of effects of stereotypes in gender-diverse work teams, as they do not only influence behavior through cognitive mechanisms—as prior research showed—but also alter our interpersonal behavior, and individual performance above and beyond ability. By shaping dominance behavior (under pro-similarity beliefs) and individual performance in stereotype-confirming ways, our study suggests that these behavioral and performance consequences of stereotypes in team contexts should receive much more attention—in theory as well as in practice. Our finding that a simple manipulation of diversity beliefs can disable the link between stereotypical perceptions and interpersonal behavior as well as performance could serve as an initial starting point for further research in the field and for practical interventions alike.

References


This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.


Received June 4, 2020
Revision received October 13, 2021
Accepted October 16, 2021