

The Sound of Support: Gendered Voice Agent as Support to Minority Teammates in Gender-Imbalanced Team

Angel Hsing-Chi Hwang
hh695@cornell.edu
Cornell University
Ithaca, New York, USA

Andrea Stevenson Won
a.s.won@cornell.edu
Cornell University
Ithaca, New York, USA

ABSTRACT

The present work explores the potential of leveraging a teamwork agent’s identity – signaled through its gendered voice – to support marginalized individuals in gender-imbalanced teams. In a mixed design experiment ($N = 178$), participants were randomly assigned to work with a female and a male voice agent in either a female-dominated or male-dominated team. Results show the presence of a same-gender voice agent is particularly beneficial to the performance of minority female members, such that they would contribute more ideas and talk more when a female agent was present. Conversely, minority male members became more talkative but were less focused on the teamwork tasks at hand when working with a male-sounding agent. The findings of the present experiment support existing literature on the effect of social presence in gender-imbalanced teams, such that gendered agents serve similar benefits as human teammates of the same gender identities. However, the effect of agents’ presence remains limited when participants have experienced severe marginalization in the past. Based on findings from the present study, we discuss relevant design implications and avenues for future research.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI; Collaborative and social computing theory, concepts and paradigms.**

KEYWORDS

Human-Agent Teamwork, Voice Agent, Minority, Majority, Marginalization

ACM Reference Format:

Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2024. The Sound of Support: Gendered Voice Agent as Support to Minority Teammates in Gender-Imbalanced Team. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 22 pages. <https://doi.org/10.1145/3613904.3642202>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0330-0/24/05

<https://doi.org/10.1145/3613904.3642202>

1 INTRODUCTION

In teamwork, we all hope for a certain degree of support. Teamwork research across the fields has consistently shown support from teammates serves as one of the most critical factors that drive positive experience and performance of individuals in teams [33, 34, 99]. However, not all teams offer the most supportive atmosphere, and working in *imbalanced teams* where there are majority and minority individuals can further hinder such favorable team outcomes. Moreover, abundant research has found minority team members are more likely to endure inferior experiences due to isolation, difficulties in building rapport, receiving less supportive feedback, and even greater pressure to act according to their identity stereotypes [2, 26, 32, 35, 39, 70, 78].

This phenomenon of unequal team experiences is particularly prominent in fields with long-standing imbalanced demographic compositions [67, 127, 139]. For example, educational research shows the existing gender gap has continuously caused discouragement for female students across STEM fields. When assigned to small learning groups for programming, male students showed significantly better peer learning outcomes than female students [27]. Similar phenomena are also found in workplaces at tech companies, and female employees – more often working as minority members on their teams – reported lower degrees of satisfaction at work [89]. Their experiences as minority members in their disciplines have made it difficult for them to find support, both professionally and socially, that helps them thrive. In response to these issues, researchers and practitioners have made abundant efforts to mitigate this challenging reality. However, limited work has explored having intelligent agents directly participate in teamwork as a form of support.

Meanwhile, extensive work in human-robot interaction (HRI) and human-agent teamwork (HAT) has consistently demonstrated positive influences of robots and agents in teams [72, 79, 80, 99]. A bot teammate can effectively support teamwork through both functional and social approaches; these include but are not limited to presenting helpful information to facilitate teamwork [88], carrying on a conversation when the team experiences bottlenecks [9], and providing emotional support through words of encouragement [109]. In particular, prior HAT work has shown that bots can encourage more equal contributions and experiences of teammates, such as through moderating turn-taking and cueing up certain teammates to balance the amount of talking across teammates [92]. Together, previous literature implies the great potential of agent teammates to create more equal and inclusive team experiences.

In the present work, we set off to tackle a particular challenge: *How can we provide computer-mediated support to minority individuals who struggle with inferior experiences and performance in*

imbalanced teams? Here, we focus on experiences of **minority** individuals – these are those who identify as the numerically smaller group within a community [68, 101, 124, 137].¹ We explore the potential of supporting minority individuals in gender-imbalanced teams by presenting either a same- or different-gender agent to participate in teamwork.

We tested this idea through a 2 (participant’s gender: female or male)² x 2 (participant’s role: minority or majority team member) x 2 (voice agent’s gender: female or male) mixed design experiment with 178 participants and observed interactions across 60 gender-imbalanced teams (each team either had more females than males, or more male members than females) and through a total of 120 teamwork sessions. Through two randomized teamwork sessions, each team has the opportunity to work with a female voice agent and a male voice agent on two different teamwork tasks.

It is worth noting while imbalanced group composition can occur in various forms and based on various personal identities (e.g., gender, ethnicity, nationality, or age), we focus on gender-based minority experiences in the present study for two major reasons. First, the issue of gender imbalance remains common in a large number of fields and requires timely remedies [65]. Second, applying gender cues to autonomous (voice) agents has become a common design practice [135]. On one hand, we see the need to understand the implications of gender assignment for agents and add to the growing literature on this topic; on the other hand, we ask whether we can leverage and turn this design trend into meaningful support. As an initial attempt to explore the use of autonomous agents for supporting female and male minority members, we adopted female- and male-sounding voices as one of the simplest ways to attach identities to autonomous agents. This allows us to explore the effect of gendered agents in teamwork without the need for further anthropomorphism, minimizing possible confounds [108].

Results of the study showed significant gender differences in participants’ self-report and behavioral responses when working with a same- or different-gender agent in teamwork. In particular, female participants showed more positive perceptions toward a same-gender agent. They also performed better (in terms of coming up with more ideas in a team brainstorming task) with the presence of a female agent. However, while male participants did not show a preference for a team agent’s gender, they performed worse when they were the only male member on a team and when a male agent facilitated the teamwork session.

The present work makes two key contributions. First, our study showed the presence of a same-gender agent in teams can effectively support minority female members, while this form of support can be attempted simply by manipulating the voice of an agent. To the best of our knowledge, our study is one of the very few studies that has explored the role of agents in facilitating minority teammates. Second, we replicate findings from prior social science research

[12, 105] and showed that females and males’ different responses and behaviors in gender-imbalanced teams also translate to human-agent teamwork. This further extends the literature on human-agent teamwork, suggesting a greater variety of human-human team dynamics might also arise in teamwork with autonomous agents.

2 BACKGROUND AND RELATED WORK

2.1 Minority Experiences in Group Settings

Individuals are inclined to identify and demonstrate different behaviors toward ingroup and outgroup members in group settings [8, 20, 30, 63, 71, 76, 128, 134], and ingroup favoritism is highly common in interpersonal settings [47, 84]. As minority members in teams (i.e., when one identifies with a group of numerically smaller size within the entire team), they often experience differently during teamwork than those of the majority group [94, 137]. With fewer ingroup fellows, minority individuals are more likely viewed and treated as outgroup by most of their teammates. Such differing experiences become more prominent when one works in a highly imbalanced group (i.e., when the sizes of majority versus minority groups are of apparent disparity) [14, 114].

In human teamwork and small-group research, scholars found altering group composition – namely, having more or fewer teammates of similar or different identities – influences minority individuals’ experience [77, 136]. However, it remains unclear whether similar effects exist in human-agent teamwork, such that placing an autonomous agent with shared or distinct cues could alter minority individuals’ experiences in teams. More concretely, the present research is motivated to address these inquiries by investigating human-agent teamwork outcomes in gender-imbalanced teams:

*RQ1: Will the presence of a **same-gender** agent influence the experience of a minority member in a gender-imbalanced team?*

*RQ2: Will the presence of a **different-gender** agent influence the experience of a minority member in a gender-imbalanced team?*

In the following, we review and ground the present work on three streams of literature: (1) The effect of having similar (ingroup) versus distinct (outgroup) teammates in teams and their influences on minority and majority individuals; (2) Prior efforts to apply technology-mediated solutions to create more equal experiences for minority and majority members; (3) Gender differences when one identifies as minority members in groups.

2.2 Presence of Ingroup and Outgroup Members in Teams

The effect of ingroup/outgroup perceptions also explains why the presence of similar members has substantial influences on teamwork. The Theory of Similarity Attraction [22] and relevant team research have validated conventional wisdom such as “birds of a feather flock together”. Individuals not only favor but also actively seek those who share similar attributes in group settings [14]. In teamwork, being able to identify and work with ingroup members increases a sense of belonging and perceived team membership

¹The concept of *minority* is distinct from *marginalization*. The former describes the fact that a group is of numerically smaller size within a society. The latter indicates a group experiences exclusion, isolation, and inferior treatment. While minority individuals do experience marginalization in a wide range of social and interpersonal settings, being in the minority group does not equal being marginalized, and vice versa.

²We made this deliberate word choice (i.e., using female/male instead of woman/man) to describe gender identities following the latest guide from American Psychology Association (<https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/>) to adopt the bias-free language.

(i.e., the extent to which a person sees themselves as a member of the team) [42]. These perceptions raise positive team experiences and motivate team contribution [34]. Indeed, people tend to contribute more work – and of higher-quality work – when they identify more with the team, leading to better team performance.

The presence of similar versus different teammates also shapes one’s expectations that influence their behaviors in teamwork [39, 78, 125]. Research in organizational behaviors has found individuals felt greater pressure to conform to identity-related stereotypes when there were fewer team members with the same identities. For instance, Johnson and Schulman found females contributed less in team decision-making when they were minority members in a team [78]; as they tended to act according to gender stereotypes in such settings, females became quieter and yielded more to male teammates. Furthermore, as individuals tend to agree with and offer more frequent support to ingroup members, they likely assume the same tendencies from teammates with similar characteristics. Therefore, people feel more encouraged to express their opinions and contribute more to team conversations when they find more teammates in common.

However, these phenomena also imply that minority individuals often face more challenges in building rapport and finding support in group settings since, to begin with, there are fewer similar members they identify as ingroup [6, 52, 103]. For instance, research studying collaboration in academia has revealed minority individuals often face more obstacles to finding collaborators and forming research teams [23, 86]. Along the same vein, longitudinal research examining gender disparity in the STEM fields has found the lack of support for female students in group settings has turned into a primary obstacle that eventually discourages many from pursuing careers in STEM [67, 127, 139].

Likewise, research on organizational behavior has also revealed the threat of imbalanced group composition and challenges commonly faced by minority group members [6, 52, 103]. Oftentimes, they experience differently from the majority members in groups as others demonstrate distinct behaviors toward ingroup versus outgroup members [25, 45, 138, 141]. For instance, in a study utilizing corporate email records, Milkman and colleagues found that individuals were less likely to respond or took longer to respond to emails from colleagues of gender or racial outgroups [102]. Numerous studies about interpersonal communication have also suggested that individuals can demonstrate distinct – and sometimes discriminatory – behaviors toward outgroup, minority individuals through both verbal and non-verbal cues [29, 50, 64, 75, 118].

All in all, minority and majority members are likely to undergo different experiences in group settings as individuals tend to treat ingroup and outgroup distinctively. In other words, minority individuals more often experience the case when their teammates interact with them as outgroup members, while majority members are more often treated as ingroup of their teammates [14, 114, 136, 148]. Conversely, research and practices in organizational behavior and management science have found that this issue can be effectively mitigated through a more balanced group composition where everyone can readily identify ingroup fellows [77]. However, this solution is not always plausible in real-world scenarios. Hiring new members and reforming teams take time and resources, while team composition often needs to prioritize individuals’ skills over other

factors. As we conceive alternative, technology-mediated solutions, we ask whether altering team composition by placing autonomous agents in teams can accomplish similar outcomes. For instance, will adding a male agent to a male-dominant group reinforce the perception of a female teammate as a minority? Alternatively, will placing a female agent on the team mitigate this experience as a minor team member? We explore these questions further in the present research.

2.3 Technological Support for Minority Individuals

Existing work in the HCI field has dedicated much effort to supporting minority individuals. A large portion of these studies focused on addressing diversity challenges in educational programs, aiming to raise minority members’ learning interests and create more inclusive environments [38, 56, 97]. Besides, a number of studies proposed new techniques that introduce changes in team formation and group dynamics in educational and professional settings, such as changing the gender ratio of small teams in a workplace [41, 61]. Prior work has also designed networking tools to connect individual workers or students, supporting minority individuals through establishing social bonds [74, 140]. Finally, several studies leveraged natural language processing to detect toxicity and discrimination in online environments, protecting minority individuals from potential discrimination [24, 66, 91].

On the other hand, adopting direct interventions in small-group interactions to support minority individuals remains relatively scarce. As one of the very few cases, [123] experimented with having agents in teams to create “pseudo presence.” For example, a sole female teammate would be assigned multiple agents to form their own “team” to de-emphasize the solo status of the minority teammate. In another example, [92] adopted a conversational agent to open up the speaking floor and interrupt certain teammates when they have taken their turn to speak a couple of times in a row. Moreover, interruptions of an agent during conversations also allow more time for non-native speakers to jump in with clarification questions and mitigate language barriers [49].

However, researchers have expressed concerns around such targeted forms of support (i.e., support that may seem to favor minority individuals), as such approaches risk putting minority members on the spot and causing greater discomfort in team experiences [48, 59, 60]. In response to this concern, we further explore how to design subtle but effective interventions that provide “just enough” support to minority individuals without causing additional awkwardness. Specifically, we examined in the present study:

RQ3: Can the presence of a gendered agent sufficiently improve team experience and performance of minority team members?

Through the literature review, we also noticed a research gap such that prior studies have shown limited consideration for individual differences in minority members’ needs and experiences during teamwork. That is, these prior studies focused on comparing minority teammates’ experiences to those of the majority members, but they overlooked individual differences *within* the minority group. Therefore, given that the present research focuses on gender-imbalanced teams, we take a closer look at how females and males

respond differently when in the role of minority members in the following section.

2.4 Gender Differences in Response to Minority Experiences

Abundant work in psychology, communication, organizational behavior, management science, and more recently, HCI, has investigated gender differences in teamwork [13, 51, 53, 55, 100, 142]. Much of the literature has examined the importance and potential benefits of gender diversity in teams, as gender imbalance remains a common phenomenon in a large number of fields. It is worth noting that, though gender is a socially constructed concept (namely, one often forms their gender identity through interacting with others in their social world) and goes beyond the female-male dichotomy [7], the majority of research in this topic area focuses on differences between self-identified males and females. Therefore, the following literature review and the study design follow the same convention. Specifically, existing literature has studied gender differences in teamwork through two closely related lenses: first, gender differences at the individual level, which drive female and male teammates' distinct behaviors in collaboration; second, how group composition (i.e., the proportion of female vs. male teammates in a team) amplify or suppress such individual differences. In the following, we further unpack these two lines of work [5, 7, 75, 93].

Prior research draws heavily on the Theory of Social Identity [133] to understand gender differences at the individual level, suggesting that gender-based behavioral differences are developed through the expectations, role models, and even stereotypes one gradually "learns" through interacting with others in society. Therefore, gender differences become particularly salient in an environment where gender-based role assignments are clearly defined – as individuals tend to develop and acquire relevant skills and the characteristics that support them to fit these expected roles [5]. As a result, research has consistently found women exhibiting a significantly higher level of social sensitivity, based on their greater ability to read nonverbal cues and make accurate inferences about what others are feeling or thinking [143]. This trait also transforms their behaviors in teams, leading female teammates to be significantly more interpersonally oriented than males in group settings [51]. Accordingly, women pay more attention to ensuring equality and inclusiveness in teams, resulting in more even turn-taking when there are more female members on a team [15, 95, 143].

By contrast, men are more likely to adopt an autocratic style, demonstrated through behaviors such as giving orders, interrupting others, and applying more assertive language [51, 112]. During teamwork, men also tend to be more task-oriented, prioritizing whether a team has achieved its goal over the experiences of individual teammates [15, 95]. Research in management science has found men exhibited a stronger tendency to claim leadership in teams. Because men expect clearer role specialization in team structures, when no one plays the leadership role, they are more likely to jump in and provide direction to the team [105]. Meanwhile, compared to females, male teammates apply a considerably distinct set of verbal and non-verbal cues in team conversations; for instance, men display more direct and dominant behaviors when speaking in teams, such as chin thrusts, gesturing, and direct eye contact,

whereas women engage in more smiling whether they are speaking or listening [46].

These behavioral differences at the individual level also cause females and males to react differently – and receive different reactions – when they encounter gender-imbalanced teams, and this becomes particularly salient when one experiences a solo status (i.e., a person is the only male or the only female on a team) [81, 82]. Female solos tend to experience greater visibility and scrutiny of their work, as well as confinement to tasks that are stereotypically feminine [18, 81, 82, 145–147], whereas male solos tend to be evaluated more positively than female solos for task performance [40, 54, 110, 119]. Empirical research has built on this prior literature and found women in male-dominant groups tend to speak less and be less assertive than men; on the contrary, male teammates in female-dominant groups tend to speak more and become even more dominant and task-oriented [12, 105]. Once again, these differences between females' and males' responses are magnified when they are the sole group members of their gender [12, 105].

Finally, these behavioral differences between female and male members can also be attributed to individuals' expectations, role entrapment, and pressure to conform to gender stereotypes [39, 78, 125]. Such tendencies become more prominent when there are fewer gender-ingroup members in a social setting. That is, when one is the sole male/female on a team, they experiences greater pressure to act according to stereotypes of males/females [78]. As a result, minority males become more talkative and more likely to interrupt others in teams when their societies view males with a higher social hierarchy; conversely, minority females become quiet due to social norms and expectations for them to yield to males [125].

Synthesizing the above literature, we hypothesize that female and male participants would respond differently when they are the sole minority members of a team. However, it remains unknown whether having a same-gender or different-gender agent on the team would provide similar support for minority male and female individuals. Namely, we examine the following inquiries in the present research:

*RQ4: How will having a **same-gender** agent on the team influence **female** (RQ4a) and **male** (RQ4b) minority members respectively?*

*RQ5: How will having a **different-gender** agent on the team influence **female** (RQ5a) and **male** (RQ5b) minority members respectively?*

3 METHOD

3.1 Overview of Study Design

We conducted a $2 \times 2 \times 2$ mixed design experiment. Participants worked on two randomized tasks in a gender-imbalanced team with an agent teammate of a certain gender identity. The between-subject variables are *participant's self-identified gender* (female or male) and *participant's assigned group* during teamwork (either as the minority or majority group). The within-subject variable is the *voice agent's assigned gender* (female or male), as all participants worked on two tasks, one with a female agent and one with a male agent in randomized orders. Note that this last variable also

defines whether the agent is of the *same or different gender* as each participant.³

3.2 Participants

We recruited participants through Prolific (N = 178) and randomly assigned them to gender-imbalanced teams (either a female-dominant or a male-dominant team) based on their self-reported gender identities. Each teamwork session included three participants; a female-dominant team consisted of two females and one male (not including the agent), while a male-dominant team had two males and one female (not including the agent). We recruited and conducted the study entirely online in the hope of getting access to participants with more diverse backgrounds and real-world work experiences. (In particular, one of the teamwork tasks required participants to evaluate whether several employees should be offered job promotions given their work profiles.) We recruited participants on Prolific because recent studies have consistently found participants on Prolific providing better and more consistent data quality than other crowdsourcing platforms, such as Amazon MTurk (See a review at [111]).

We ran 20 groups of participants as a pilot to ensure the validity of the study protocol.⁴ As the pilot already showed significant patterns with the behavioral data, instead of performing power analysis, we followed the convention in human subject research [85] and set the target sample size as 180 such that each condition at least includes 30 participants (i.e., having 30 females as minority members, 30 males as minority members, 60 females as majority members, and 60 males as majority members). According to the Central Limit Theorem in the statistics literature, it requires at least $n = 30$ for a sample to reach normal distribution, which is the statistical assumption for various tests performed in the current data analyses [132]. Given this target sample size (180 in total), we set off to recruit 90 females and 90 males. In the end, data from 2 participants were removed because they did not complete the post-teamwork survey, resulting in N = 178 as the final sample size. Participants' demographic data is recorded and further elaborated in the Measurement section. Among the 178 participants used for final data analysis, there are 90 females and 88 males. These include 30 female-dominant groups and 30 male-dominant groups.

3.3 Procedure

Upon signing up to take part in the study, participants first filled out a short pre-survey to self-report their past experiences of teamwork and marginalization, as well as their self-identified demographic profiles and familiarity with AI technologies. Given the design of our study and the voice agent (i.e., participants could only differentiate the agent by having either a female or male cue), we did not include participants who did not identify themselves as female/male. As such, one crowdsourced worker was compensated for filling out the pre-survey but did not proceed with the study. We acknowledge and further address the limitation of this binary approach in Discussion.

³To check whether our findings indeed result from participants working with either a same- or different-gender agent – instead of due to the gender of the voice agent per se – we performed an analysis in Section 4.2.2 to rule out this alternative explanation.

⁴We ran these 20 groups solely for pilot purposes to ensure the study protocol worked well. Data from the pilot were not included in the final analyses.

Participants were then randomly assigned to one of the two types of gender-imbalanced teams and attended a teamwork session based on their availability. All teamwork sessions were held remotely on Zoom. In each of the teamwork sessions, participants were asked to work on two teamworking tasks in a randomized order (an idea generation task [120] and a decision-making task; see more details in the "Teamwork Tasks" section). We randomly assigned the gender identity of the agent, such that in each session, participants worked with a female agent in one task and a male agent in the other task, but the order in which the female agent and the male agent were presented was random. This experimental design resulted in four possible task procedures:

- (1) In the first task, participants worked with a **female** agent on an idea generation task. In the second task, participants worked with a **male** agent on a decision-making task.
- (2) In the first task, participants worked with a **male** agent on an idea generation task. In the second task, participants worked with a **female** agent on a decision-making task.
- (3) In the first task, participants worked with a **female** agent on a decision-making task. In the second task, participants worked with a **male** agent on an idea generation task.
- (4) In the first task, participants worked with a **male** agent on a decision-making task. In the second task, participants worked with a **female** agent on an idea generation task.

Immediately after completing each task, participants were asked to fill out a short survey to reflect on their teamwork experiences and their perceptions toward their teammates. After they completed both tasks and filled out both surveys, participants completed a short exit survey to share their overall experiences and final thoughts. As we are most interested in minority individuals' experience during this human-agent teamwork setting, we spent an additional 15 - 20 minutes interviewing participants who were assigned as minority members after they completed the experiment and surveys to collect their qualitative feedback. The entire study took around 45 ~ 60 minutes to complete, and participants received \$15 cash compensation for their time and work. The study protocol was reviewed and approved by the Institutional Review Board (IRB).

3.4 Teamwork Tasks

Participants worked on two teamwork tasks in a randomized order. One task was an idea generation task where participants were asked to come up with ideas for water and/or energy conservation with their teammates [120]. Participants had 10 minutes to work on this task. The task has been used in several prior teamwork studies (e.g., [129]), as the topic requires ideas to be not only new but also practical, meeting the definition of creativity as "novel and functional" [126].

The other task was a decision-making task where participants viewed the profiles of five employees and determined who should be offered a promotion. The task and these employee profiles were adopted from Binns et al.'s oft-cited study [16], which has been used in prior research studying human-AI co-decision-making [87]. To ensure participants came to a decision for each employee profile, the AI agent explicitly asked participants to vote on whether they would promote each candidate or not before moving on to the next one. Participants had 15 minutes to work on this task. The lengths of

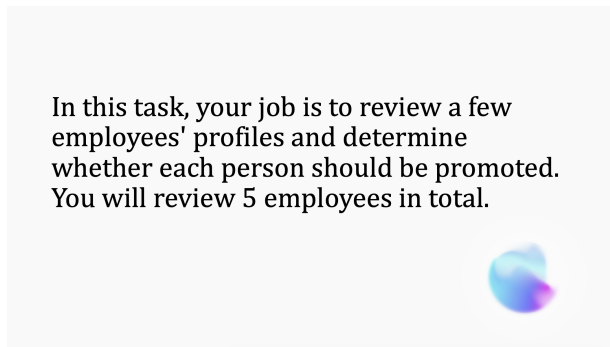


Figure 1: A screen capture of the interface used for the present study.

both teamwork tasks were determined by conducting pilot studies with other research assistants.

During both tasks, participants saw a screen presenting the teamwork task content (see Figure 1). They interacted with their human teammates and the AI agent merely through voice interaction. The AI agent was not embodied in a physical form or humanoid figure. The only visual cue of the AI agent was an animated, abstract shape shown on the screen. We deliberately chose to minimize the tangible form and visual representation of the AI in order to eliminate the confounding factors of embodiment and anthropomorphism [108].

3.5 AI Agent in Teamwork

The AI agent in the present study was implemented through a Wizard-of-Oz protocol. The agent's responses were pre-scripted and pre-recorded. We presented the scripts we used in Appendix. The first author played the agent in each teamwork session by playing the pre-scripted content, which was all generated by OpenAI's ChatGPT⁵. The decision to adopt a Wizard-of-Oz paradigm instead of using an existing voice assistant product (e.g., Apple Siri, Amazon Alexa, or Google Assistant) was made deliberately so that the agent's speech content in each team conversation was identical. When participants' responses could not be covered using the pre-recorded content, the agent would respond with "I'm sorry, but I don't understand what you just said."

During each teamwork task, the agent served the following functions and participated in team discussions. The experimenter did not intervene during the teamwork sessions at all.

- (1) At the beginning of each task, the agent greeted and read out the study instructions on screen. It also signaled when participants could start their discussion.
- (2) For every three turns taken by participants, the agent would jump in once to participate in the team discussion. In the idea generation task, the agent contributed one idea each turn. In the decision-making task, the agent provided one argument for whether the candidate should be promoted each turn. On average, the agent took 7.9 turns during each teamwork task.
- (3) The agent handled time-keeping. It would let participants know when they hit the time limit. Immediately afterward,

⁵<https://chat.openai.com/>

the experimenter would show up to direct participants to complete the post-survey.

The key manipulation in the present study was the gender identity of the AI agent. We used the agent's voice (i.e., either a female-sounding or male-sounding voice) to implement such manipulation. After obtaining the speech content from ChatGPT, we then created the female- and male-sounding responses through Google Cloud's text-to-speech function. It is worth noting that the sound quality of the voice agent is not the focus of this study. Instead, what matters is the gender identity that participants associate with the agent. Existing work has found that users tend to categorize an agent's voice as female-sounding or male-sounding [113]. Therefore, we did not include a "gender neutral" condition because adopting a gender-ambiguous voice for the agent may simply add cognitive load to participants as they strive to identify the agent as female or male.

3.6 Measurement

All teamwork sessions were video recorded through Zoom and, later on, transcribed through Otter.ai⁶. These materials served as the main data source for analyzing participants' behaviors during the teamwork sessions. Additionally, we captured a wide range of self-report responses from participants through the pre- and post-task surveys. Measurement in the present study can be categorized into the following four themes.

3.6.1 Behavioral responses. Transcripts of participants' conversations during teamwork were labeled with their subject IDs and timecode. We first examined the behavioral data by looking at the *portion of speech* of each participant in their team ($M = 0.27$, $S.D. = 0.16$). Namely, this measure captures how much time each participant talked during a team conversation divided by the total amount of time all members of the same team spoke in the conversation.⁷ For the idea generation task, we further coded the *number of ideas* generated by each participant ($M = 3.10$, $S.D. = 2.28$). The coding result was verified by a second coder going through the transcripts to code the number of ideas generated as well. The inter-rater reliability reaches 89.99%. For any discrepancy, we took the mean value between the two coding results.

3.6.2 Teamwork outcomes. In the brainstorming task, participants' team performance was captured by the number of ideas generated. In each brainstorming session, each team produced 10.11 ideas on average ($S.D. = 4.86$). In the decision-making task, participants were asked to vote on whether they would promote an employee or not after reviewing each profile. While there was no "right" answer to the decision-making tasks, we looked at whether participants formed consensus within their teams. After voting, participants

⁶<https://otter.ai/>

⁷We use speech time instead of word count to measure each participant's portion of speech because using a premium transcribing tool (i.e., Otter.ai Premium), participants' recorded speech was still not 100% transcribed word by word. We noticed missing words in transcriptions more likely occurred when multiple participants spoke at the same time, while the time code of the audio recording remained accurate for each participant even when multiple people spoke at the same time. We checked the correlation between word count and time and saw a high correlation between the two variables ($\rho > 0.85$). We therefore proceed with speech time as a more accurate measure.

were able to reach a consensus among 84.84% of all decisions discussed during the teamwork conversations (i.e., when all teammates agreed to promote or not to promote an employee). These results were again reviewed by two coders reading through the recording transcripts. As these behavioral data were not measured on fixed scales (e.g., unlike self-report measures using 7-point Likert scales), we used their standardized values for data analyses to prevent results from being biased by extreme values.

3.6.3 Self-report responses. Participants self-reported measures in both the pre- and post-study questionnaires. All items were measured on a 7-point Likert scale unless specified otherwise (1 = strongly disagree; 7 = strongly agree).

To understand how individuals' prior experience of marginalization may influence their current team behaviors and the effect of having an agent teammate, we asked participants to report their existing experience of *marginalization* in the pre-survey through three scales from existing literature [31] ($M = 2.85$, $S.D. = 1.16$). These scales capture key dimensions of marginalized experience, including whether individuals have felt a sense of ignorance in group settings (e.g., feeling "invisible" or "unheard") and whether their peers have demonstrated micro-aggressive behaviors against their identities.

During the study, participants filled out two post-surveys, each taking place immediately after finishing each teamwork task. The post-survey captured whether participants experienced *marginalization* during teamwork using the same items as in the pre-study questionnaire ($M = 2.17$, $S.D. = 1.99$). They reflected on whether they found themselves being marginalized by their human teammates ($M = 2.47$, $S.D. = 2.03$) or by their AI teammates respectively ($M = 2.64$, $S.D. = 2.05$). Furthermore, participants reported their overall *team experience* ($M = 6.07$, $S.D. = 0.79$), with higher values indicating more positive team experience [19]. Besides, participants reported their *perception toward human teammates* ($M = 6.20$, $S.D. = 0.78$) and *perception toward AI teammates* respectively ($M = 4.75$, $S.D. = 1.49$) wherein higher values indicate a more positive perception of their teammates [92]. Participants also specified their perceived *team support* either from their human teammates ($M = 5.13$, $S.D. = 0.68$) or from their AI teammates ($M = 4.40$, $S.D. = 0.95$) [37]. Finally, we adopted six items from the NASA task load scale [69] to measure participants' perceived workload and difficulties of the teamwork tasks ($M = 2.67$, $S.D. = 0.98$).

3.6.4 Demographic data. In the pre-survey, participants were asked to self-report their basic demographic profiles, including age, gender, ethnicity, years of work experience, and fluency in the English language (since the entire study, including the team interaction and the questionnaires, was conducted in English). Based on participants' self-report profiles, there are 90 females and 88 males in the current sample. Their average age is 34.90 ($S.D. = 11.39$). The most identified ethnicity groups are Caucasian (40.82%), Black American (17.75%), and Asian (16.57%). On average, participants have 10.98 years of work experience ($S.D. = 11.17$). 70.69% of participants mentioned English as their first language, while all of them were able to communicate fluently in English during the team conversations. All except for 6 participants have used a voice assistant before, and 23.60% of participants reported themselves as regular users.

3.6.5 Validity check. At the end of the second post-survey (i.e., the final questionnaire participants filled out after completing both teamwork tasks), we included an open-ended question for a validity check, asking participants to describe what they thought the study was about in text. This question is used to ensure participants were not fully aware of the experimental manipulation. In their responses, none of the participants mentioned team marginalization as the topic of the study. Furthermore, to check the effectiveness of the Wizard-of-Oz study protocol, we asked participants whether they believed the AI agent was indeed an AI or was played by a human. Again, all participants believed the teamwork agents were indeed AI instead of humans.

4 RESULTS

4.1 Overview of Data Analysis

We performed all data analyses in R [115]. We began the analyses by examining whether participants' self-report and behavioral responses differed by (1) their gender, (2) whether they were minority or majority members, and (3) whether the team agent adopted a female or male voice. For each independent variable, we first ran a factorial ANOVA to check whether there was a significant mean difference. For independent variables that showed significant mean differences, we then used linear models to examine the relationships between variables. Depending on whether a dependent variable is a repeated measure or not, we used either the `lm` or `lmer` function in R to perform linear modeling. For repeated measures (i.e., self-report scales in the two post-task questionnaires), we used linear mixed models from the `lmer` package [11] to account for within-subject variances by including the (1 | subjectID) term in the models. All linear models (including both `lm` and `lmer` models) were controlled for the task order (i.e., whether the data was collected from the first or second teamwork task in a study) to account for a potential order effect resulting from the between-within-subject design.

Then, we examined whether female and male participants responded differently in minority situations by testing a two-way interaction term of *participant's gender* (female/male) \times *minority teammate* (yes/no) in linear models. Likewise, we are interested in whether the presence of a same- or different-gender agent affects female and male minority teammates' experience and performance in teamwork. We performed this part of the analysis by coding each participant's assigned team scenario into three types. Participants who were randomly assigned as the majority group were labeled as receiving *human support*. Participants who were randomly assigned as minority members were coded as receiving either *AI support* or *no support*. To be specific, *AI support* indicates when a minority teammate worked with a same-gender agent; *no support* indicates when a minority teammate worked with a different-gender agent.

After coding these three support types, we again tested whether the support types have any significant main effect on participants' self-report experience and behavioral responses. Furthermore, we examined whether there is a significant two-way interaction between participants' gender and the support types they received. We adopted this analytic approach instead of testing a three-way interaction of participant's gender (female/male) \times minority teammate (yes/no) \times agent's gender (female/male) as the former can

Purpose	Model	Independent Variable (IV)	Dependent Variable (DV)	Moderating Variable (MV)
[Baseline] To examine how female and male participants responded when assigned as minority team members based on their self-report scales (repeated measures) → Results reported in Section 4.2	lmer(DV ~ IV + task_order + (1 subjectID))	participant's gender × majority/minority condition	1. team marginalization 2. team experience 3. support from teammate 4. perception of teammate 5. task load	n.a.
[Baseline] To examine how female and male participants responded when assigned as minority team members based on their behavioral data (non-repeated measures) → Results reported in Section 4.2	lm(DV ~ IV + task_order)	participant's gender × majority/minority condition	1. portion of speech 2. # of ideas generated 3. agreement among teammates	n.a.
[Baseline, RQ1, RQ2] To examine the main effect of experimental conditions on self-report scales (repeated measures) → Results reported in Sections 4.2 & 4.3	lmer(DV ~ IV + task_order + (1 subjectID))	1. participant's gender 2. majority/minority teammate 3. agent's gender 4. presence of same/different-gender agent	1. team marginalization 2. team experience 3. support from teammate 4. perception of teammate 5. task load	n.a.
[Baseline, RQ3] To examine the main effect of experimental conditions on behavioral measures (non-repeated measures) → Results reported in Sections 4.2 & 4.4	lm(DV ~ IV + task_order)	1. participant's gender 2. majority/minority teammate 3. agent's gender 4. presence of same/different-gender agent	1. portion of speech 2. # of ideas generated 3. agreement among teammates	n.a.
[RQ4, RQ5] To examine whether and how having same- or different-gender agents has an effect on minority teammates based on their self-report scales (repeated measures) → Results reported in Section 4.5	lmer(DV ~ IV + task_order + (1 subjectID))	participant's gender × support type (human, same-gender agent, different-gender agent)	1. team marginalization 2. team experience 3. support from teammate 4. perception of teammate 5. task load	n.a.
[RQ4, RQ5] To examine whether and how having same- or different-gender agents has an effect on minority teammates based on their behavioral scales (non-repeated measures) → Results reported in Section 4.5	lm(DV ~ IV + task_order)	participant's gender × support type (human, same-gender agent, different-gender agent)	1. portion of speech 2. # of ideas generated 3. agreement among teammates	n.a.
[Effect of prior experience] To examine whether and how existing team experience moderates minority teammates' self-report experiences (repeated measures) → Results reported in Section 4.6	lmer(DV ~ IV*MV + task_order + (1 subjectID))	participant's gender × support type (human, same-gender agent, different-gender agent)	1. team marginalization 2. team experience 3. support from teammate 4. perception of teammate 5. task load	prior experience of marginalization
[Effect of prior experience] To examine whether and how existing team experience moderates minority teammates' behavioral responses (non-repeated measures) → Results reported in Section 4.6	lm(DV ~ IV*MV + task_order)	participant's gender × support type (human, same-gender agent, different-gender agent)	1. portion of speech 2. # of ideas generated 3. agreement among teammates	prior experience of marginalization

Table 1: Summary of models tested in the present study

help with interpreting the data more handily. For instance, with a three-way interaction test, looking at the effect of minority females or minority males working with a female agent has different syntactic meanings, making interpretations of the findings more challenging.

Finally, we examined whether participants' prior team and marginalization experiences would affect their current human-agent teamwork experiences and behaviors. We did so by including their self-report responses in the pre-survey as a moderating variable in the linear models and testing a three-way interaction effect across participants' gender, their prior marginalization experience, and the support types they received during teamwork. In Table 1, we summarize all models tested for the present study.

4.2 Understanding Baseline: Human-Agent Teamwork in Gender-Imbalanced Teams

4.2.1 Gender-imbalanced team composition influences individuals' perceived marginalization. A significant main effect of minoritized conditions on participants' self-report marginalized experience verified successful experimental manipulation, such that participants who were randomly assigned as the minority member in a team indeed reported a significantly higher degree of marginalization

post-teamwork ($\beta = 0.72$, $S.E. = 0.15$, $t = 4.96$, $p < 0.001$). Furthermore, participants reported marginally worse perceptions of their human teammates when they were the minority team members ($\beta = -0.22$, $S.E. = 0.13$, $t = -1.73$, $p = 0.085$, partial $\eta^2 = 0.007$).

4.2.2 Voice agent's gender does not directly influence team experience or performance. Test results of linear models did not show a significant main effect of the teamwork agent's assigned identity (i.e., presented in either a male-sounding or female-sounding voice) on any of the self-report or behavioral variables. With this null effect, we confirm the following findings have to do with whether participants worked with a same-gender or different-gender agent, rather than due to the assigned gender of the voice agents per se.

4.2.3 Female and male participants show gender differences in team experience and performance. Regardless of their assigned conditions, participants' self-report and behavioral responses showed significant gender-based differences. To begin with, male participants reported significantly more positive perceptions toward their human teammates ($\beta = 0.26$, $S.E. = 0.12$, $t = 2.19$, $p = 0.029$) and also marginally more favorable team experience ($\beta = 0.22$, $S.E. = 0.13$, $t = 1.67$, $p = 0.097$, partial $\eta^2 = 0.007$). Meanwhile, female participants reported a marginally higher level of perceived task load during teamwork ($\beta = 0.28$, $S.E. = 0.16$, $t = 1.74$, $p = 0.083$, partial

$\eta^2 = 0.02$). Demonstrated in their behavioral data, male participants talked significantly more in team conversations (statistical test results based on transformed data: $\beta = 0.07$, $S.E. = 0.01$, $t = 10.54$, $p < 0.001$; statistical test results based on raw data: $\beta = 0.47$, $S.E. = 0.04$, $t = 10.54$, $p < 0.001$). Regarding their performance in the idea generation task, male participants also contributed significantly more ideas than female teammates in the brainstorming task (statistical test results based on transformed data: $\beta = 0.07$, $S.E. = 0.04$, $t = 1.71$, $p = 0.031$; statistical test results based on raw data: $\beta = 0.02$, $S.E. = 0.01$, $t = 1.71$, $p = 0.003$). Regarding outcomes of the decision-making task, there was no significant difference in male and female participants' likelihood of forming consensus during team discussions.

Furthermore, we found female and male participants showed several different patterns when they were the minority members on their teams; that is, results of linear models showed a significant

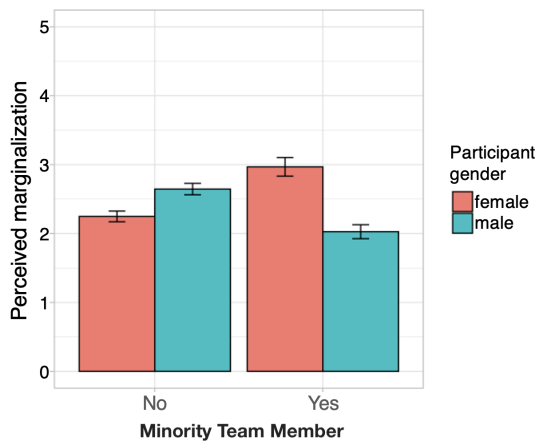


Figure 2: Self-report marginalized experience by participants' assigned condition and their gender. Whiskers represent standard errors.

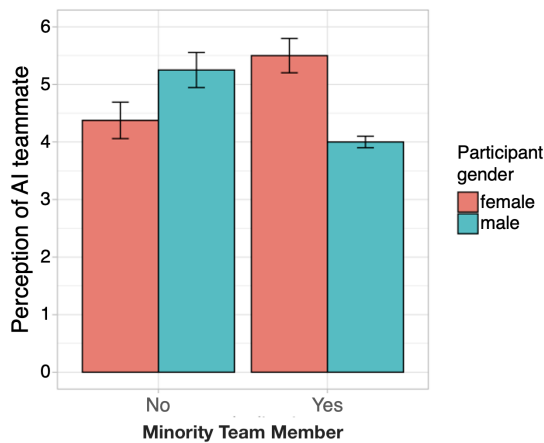


Figure 3: Self-report perception of AI teammates by participants' assigned condition and their gender. Whiskers represent standard errors.

interaction effect between participants' gender and their assigned conditions (i.e., whether one was a minority or majority member in teamwork) on their subjective experience and behavior during teamwork. We first found a significant gender difference in participants' self-report marginalization ($\beta = -1.36$, $S.E. = 0.21$, $t = -6.59$, $p < 0.001$); as shown in Figure 2, female participants reported a significantly higher degree of marginalized experience when they were the minority member in a team, whereas male participants were less likely to self-report being marginalized when they were the only male member in a team. Meanwhile, when being the minority members, male participants found the teamwork more burdening and reported a significantly higher degree of task load compared to scenarios when they were the majority group in teams; however, female participants did not reveal a similar pattern when they were the minority members ($\beta = 0.75$, $S.E. = 0.35$, $t = 2.17$, $p = 0.032$).

Gender-imbalanced group composition can also affect one's perception of their teammates. When being the minority teammates, female and male participants as well showed different perceptions toward their AI teammates ($\beta = -2.38$, $S.E. = 1.07$, $t = -2.21$, $p = 0.036$). To be specific, female participants expressed more positive sentiment toward their AI teammates when they were the only female on the team, whereas male participants reported worse perceptions of AI teammates when they the minority members in teamwork (see Figure 3).

Analyzing data from participants' team conversations, we found female and male participants also showed different behavioral patterns when they were the minority members in teams. In particular, there is a significant interaction effect between participants' gender and their assigned condition (i.e., whether they were assigned to the majority or the minority group) on the portion they spoke up during team conversations (statistical test results based on transformed data: $\beta = 0.08$, $S.E. = 0.01$, $t = 5.76$, $p < 0.001$; statistical test results based on raw data: $\beta = 0.54$, $S.E. = 0.09$, $t = 5.76$, $p < 0.001$). When being the minority teammates, female participants tended to become less active during team discussions, but male participants instead talked more when they were the only male members of a team. However, talking more did not indicate greater team contribution or productivity. In fact, though male participants talked more when they were the minority members, they generated fewer ideas compared to their male counterparts who were assigned as the majority members. Conversely, though minority female participants were relatively quiet, they were able to focus on the task at hand and generated more ideas than females who were in the majority groups. As a result, we also saw a significant interaction effect between participants' gender and assigned minority/majority conditions on the count of ideas they produced (statistical test results based on transformed data: $\beta = -0.22$, $S.E. = 0.09$, $t = -2.29$, $p = 0.022$; statistical test results based on raw data: $\beta = -0.07$, $S.E. = 0.03$, $t = -2.29$, $p = 0.022$). Together, gender-based behavioral differences in the minority versus majority conditions were illustrated in Figure 4. Meanwhile, there was no significant difference in the outcomes of the team decision-making task.

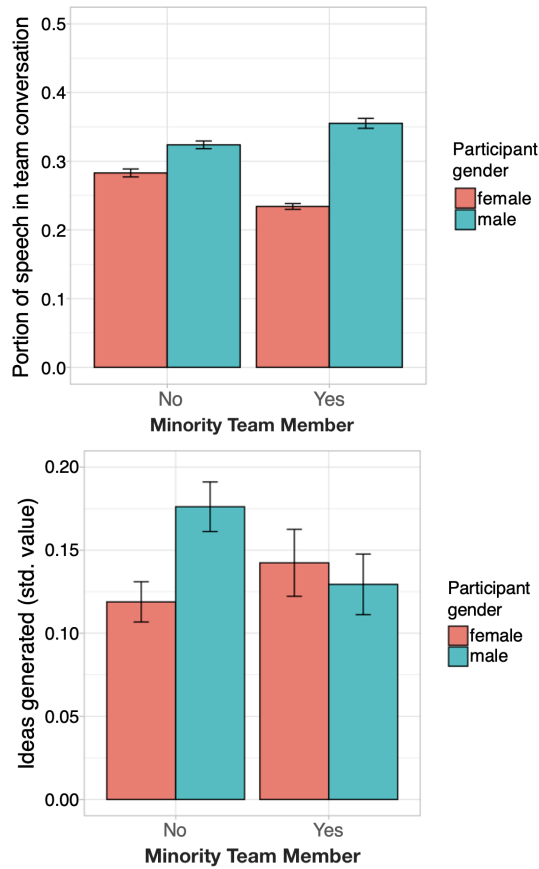


Figure 4: Behavioral difference in team conversations by participants' assigned condition and their gender. Whiskers represent standard errors.

4.3 Teamwork Experience with the Presence of A Same/Different-Gender Agent [RQ1, RQ2]

Across all participants, whether they worked with a same- or different-gender agent does not affect their teamwork experience. Test results of linear models did not show any significant difference in participants' self-report team experience ($\beta = 0.04$, $S.E. = 0.30$, $t = 0.12$, $p = 0.907$), their perceptions toward AI teammates ($\beta = 0.03$, $S.E. = 0.31$, $t = 0.10$, $p = 0.921$), their perceptions toward human teammates ($\beta = 0.04$, $S.E. = 0.31$, $t = 0.15$, $p = 0.880$), or perceived task load ($\beta = 0.02$, $S.E. = 0.16$, $t = 0.13$, $p = 0.899$).

Even when we accounted for whether a participant was a minority or majority member in their team (i.e., we tested whether there was a significant interaction effect between the presence of a same-/different-gender agent vs. participants' roles (minority or majority) in their team), there was no significant effect of the presence of a same- or different-gender agent on their team experience ($\beta = 0.09$, $S.E. = 0.65$, $t = 0.13$, $p = 0.895$), their perception toward human teammates ($\beta = 0.04$, $S.E. = 0.66$, $t = 0.06$, $p = 0.953$) and AI teammates ($\beta = 0.17$, $S.E. = 0.67$, $t = 0.25$, $p = 0.805$), or their perceived task load ($\beta = 0.11$, $S.E. = 0.36$, $t = 0.30$, $p = 0.766$).

4.4 Teamwork Performance with the Presence of A Same/Different-Gender Agent [RQ3]

Similarly, whether one worked with a same- or different-gender agent does not affect their teamwork performance. There was no significant effect of the presence of a same-/different-gender agent on the portion of team discussions one participated in (statistical test results based on transformed data: $\beta = 0.004$, $S.E. = 0.02$, $t = 0.18$, $p = 0.857$; statistical test results based on raw data: $\beta = 0.03$, $S.E. = 0.02$, $t = 1.16$, $p = 0.318$) nor on the number of ideas they contributed during teamwork (statistical test results based on transformed data: $\beta = -0.03$, $S.E. = 0.38$, $t = 0.01$, $p = 0.999$; statistical test results based on raw data: $\beta = -0.05$, $S.E. = 0.38$, $t = 0.38$, $p = 0.248$).

Again, the null effect remains even when we accounted for participants' roles in their teams (i.e., whether they were minority or majority members in teamwork). As such, there was no significant interaction effect between the presence of a same-/different-gender agent vs. the role of participants in teams on the portion of speech one participated in during team discussions (statistical test results based on transformed data: $\beta = 0.001$, $S.E. = 0.05$, $t = 0.02$, $p = 0.986$; statistical test results based on raw data: $\beta = 0.03$, $S.E. = 0.04$, $t = 0.89$, $p = 0.373$) nor on the number of ideas they contributed during teamwork (statistical test results based on transformed data: $\beta = -0.12$, $S.E. = 0.82$, $t = 0.01$, $p = 0.998$; statistical test results based on raw data: $\beta = -0.05$, $S.E. = 0.58$, $t = 0.01$, $p = 0.318$).

4.5 Gender Difference in Response to the Presence of A Same/Different-Gender Agent [RQ4, RQ5]

Participants' responses to working with a same- or different-gender agent showed significant gender differences. To begin with, there was a significant interaction effect between participants' gender and the types of team support they received on their perception toward the teamwork agent ($\beta = 3.06$, $S.E. = 1.44$, $t = 2.13$, $p = 0.044$). In particular, female participants reported significantly more positive perceptions of the AI teammate when they were the minority members – with the highest average ratings for same-gender agents. Conversely, male participants showed a worse perception of the AI teammate when they were the minority members, and they also did not have a preference between the female agent and the male agent. Compared to participants' perception of their human teammates, we did not observe significant gender difference when working with a same- or different-gender agent ($\beta = 0.09$, $S.E. = 0.41$, $t = 0.23$, $p = 0.818$). Gender difference in subjective experience during teamwork was also reflected in participants' perceived task load during teamwork ($\beta = -1.11$, $S.E. = 0.44$, $t = -2.49$, $p = 0.014$). As per the analysis from the previous section, while male teammates perceived a greater task load when they were in the minority condition, they felt more burdened when working with a different-gender agent than working with a same-gender agent. Resonating with results from the previous section, females were, by contrast, able to focus better on the task at hand when they were the minority members, resulting in a lower perceived task load.

Similarly, females and males demonstrated distinct behavioral responses to same- and different-gender agents when they were

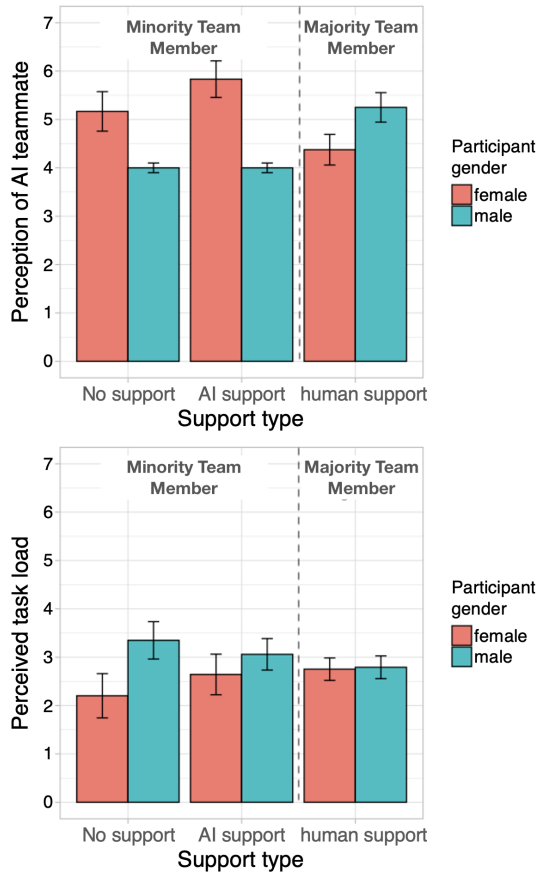


Figure 5: Effect of agent's support on participants' subjective experiences in teamwork. Whiskers represent standard errors.

minority members. First, there is a significant interaction effect between participants' gender and received support types on the portion of speech they participated in during team conversations (statistical test results based on transformed data: $\beta = -0.09$, $S.E. = 0.02$, $t = -5.32$, $p < 0.001$; statistical test results based on raw data: $\beta = -0.47$, $S.E. = 0.12$, $t = -3.90$, $p < 0.001$). Once again, male participants talked more when they were minority members, and they were particularly talkative when working with a different-gender agent in the minority condition. Conversely, while female participants talked less in a team where they were the only female member, the difference between working with a same- or different-gender agent was less prominent. Again, how much a participant talked in team conversations did not necessarily transfer to their contributions. Upon observing a significant interaction effect between participants' gender and their support types on the number of ideas they generated during teamwork (statistical test results based on transformed data: $\beta = -0.55$, $S.E. = 0.16$, $t = -3.51$, $p < 0.001$; statistical test results based on raw data: $\beta = -0.18$, $S.E. = 0.05$, $t = -3.51$, $p < 0.001$), we saw minority females thrived with the support of a same-gender agent, producing even more ideas than those females working in a female-dominant group. While minority

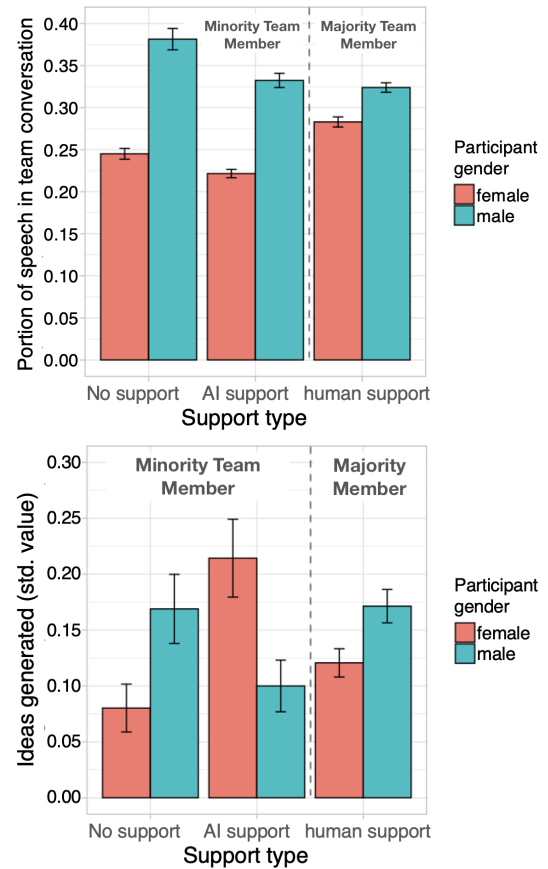


Figure 6: Effect of agent's support on participants' behaviors in teamwork. Whiskers represent standard errors.

males did not show a preference for either a same- or different-gender agent (see results reported in the above paragraph), they were able to generate the most ideas without the presence of any same-gender entity. Once again, we did not observe significant results regarding participants' tendency to form consensus during the team decision-making task.

4.6 Influences of Individuals' Prior Teamwork Experience

In the following sections, we examined whether participants' prior experience in teamwork had any effect on how they performed in the present teamwork. We began with examining the effect of whether participants have experienced marginalization in the past by testing three-way interaction terms: (1) *gender × prior marginalized experience*: a two-way interaction between participants' gender and their self-report prior marginalized experience, (2) *gender × minority/majority condition × prior marginalized experience*: a three-way interaction across participants' gender, whether they were assigned as the minority member in teamwork during the present study, and whether they have experienced marginalization in the past, and (3) *gender × support type × prior marginalized experience*: a three-way interaction across participants' gender, the type of

support they received during the present study (from a human, same-gender agent, or different-gender agent), and whether they have experienced marginalization in the past.

Before examining the above test results, it is worth noting that participants' self-report prior experience of marginalization is not normally distributed (based on a Shapiro-Wilk normality test: $W = 0.96$, $p < 0.001$), and the majority of participants did not report experiencing intense marginalization in the past (see Figure 7). Accordingly, we used the log-transformed values of this variable for analyses. However, findings concerning those who have experienced more marginalization in the past may still be subject to greater variances.

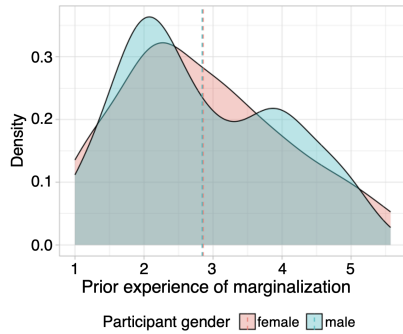


Figure 7: Density plot of participants' prior experience of marginalization. Whiskers represent standard errors.

Results of the above-mentioned analysis first showed gender differences in the moderating effect of one's prior marginalization on team behavior. First, we observed a significant interaction effect between participants' gender and their ratings of prior marginalized experience on the portion of the speech in team conversations during the current study (statistical test results based on transformed data: $\beta = 0.05$, $S.E. = 0.02$, $t = 2.25$, $p = 0.026$; statistical test results based on raw data: $\beta = 0.05$, $S.E. = 0.17$, $t = 0.27$, $p = 0.024$). As shown in Figure 8, female participants who experienced more severe marginalization in the past tended to talk less during teamwork sessions in the present study, while male participants who experienced more marginalization in the past were more talkative in the study sessions. Looking at the number of ideas generated during the present teamwork experiment, participants' prior experience of marginalization also moderated their performance in teams (statistical test results based on transformed data: $\beta = 0.64$, $S.E. = 0.31$, $t = 2.03$, $p = 0.042$; statistical test results based on raw data: $\beta = 0.21$, $S.E. = 0.10$, $t = 2.03$, $p = 0.042$). As shown in the downward-trending line on the right of Figure 8, while, in general, participants who have experienced more severe marginalization in the past tended to contribute fewer ideas during the brainstorming sessions, this trend was more salient among female participants. On the other hand, there was no significant result regarding the outcomes of the decision-making task.

Next, we examined whether and how past marginalization experience may influence female and male participants who worked either as the majority or minority during the current teamwork study. Results showed a marginally significant three interaction

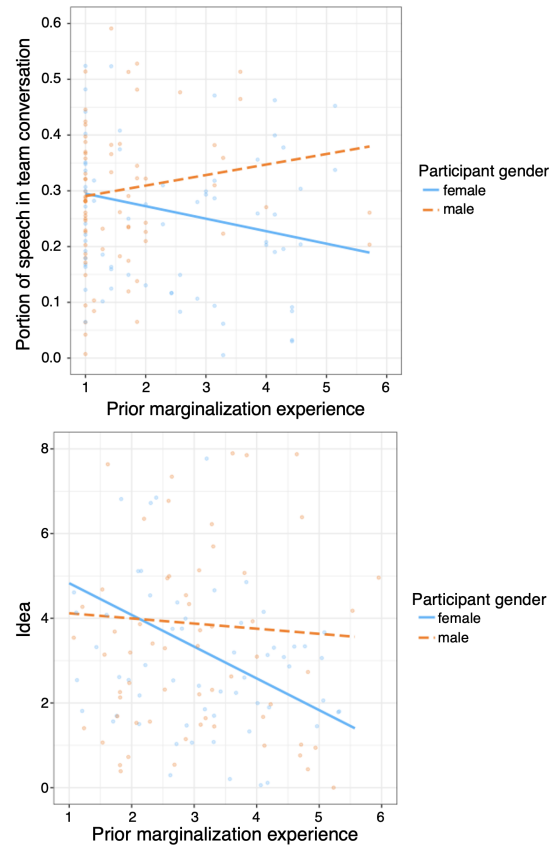


Figure 8: Moderating effect of participants' prior marginalization experience on team behaviors.

across participants' gender, their assignment as either majority or minority members, and their prior experience of marginalization on the portion of which they talked during team conversations (statistical test results based on transformed data: $\beta = -0.09$, $S.E. = 0.05$, $t = -1.87$, $p = 0.063$, partial $\eta^2 = 0.02$; statistical test results based on raw data: $\beta = -0.15$, $S.E. = 0.12$, $t = -1.16$, $p = 0.024$, partial $\eta^2 = 0.008$) as well as the number of ideas generated during the team brainstorming session (statistical test results based on transformed data: $\beta = 1.88$, $S.E. = 0.81$, $t = 2.32$, $p = 0.022$; statistical test results based on raw data: $\beta = 0.58$, $S.E. = 0.08$, $t = 0.71$, $p = 0.048$). Specifically, female and male participants who experienced more marginalization in the past contributed fewer ideas when assigned to the majority groups. However, when they were indeed placed in the minority situation, male participants who experienced more intense marginalization in the past contributed more ideas during the teamwork sessions. In contrast, females who have been marginalized in the past tended to contribute less to the minority condition during the present study as well. Again, there was no significant result regarding the outcomes of the team decision-making task.

Finally, there was a marginally significant three-way interaction effect across participants' prior marginalization experience, their gender, and the types of support they received on the number of ideas generated during teamwork (statistical test results based on

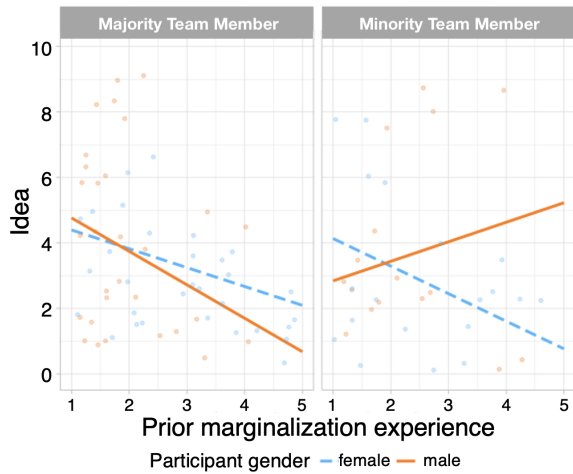


Figure 9: Moderating effect of participants' prior marginalization experience on their responses to the assignment as minority or majority members.

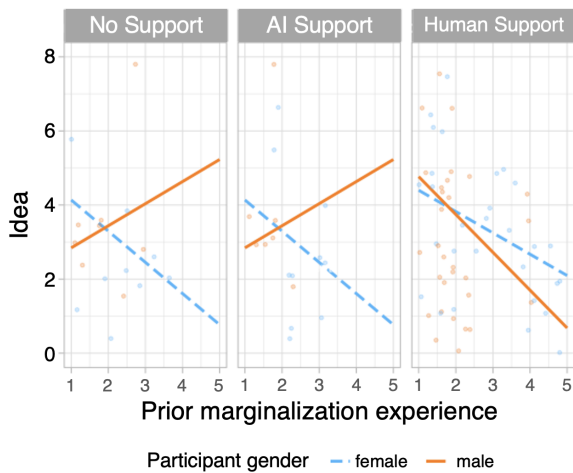


Figure 10: Moderating effect of participants' prior marginalization experience on the effect of support types.

transformed data: $\beta = -1.88$, $S.E. = 1.09$, $t = -1.73$, $p = 0.086$, partial $\eta^2 = 0.02$; statistical test results based on raw data: $\beta = -0.60$, $S.E. = 0.13$, $t = -0.46$, $p = 0.045$, partial $\eta^2 = 0.01$). Specifically, when considering participants' prior experiences, male participants who have gone through more intense marginalization benefited more from having an autonomous agent on the team, regardless of the agent's gender. As shown in Figure 10, male participants generated more ideas with the presence of a teamwork agent when they were indeed marginalized in the present experiment. Conversely, female participants who have been marginalized in the past did not show a significant difference in their work performance (evaluated by the number of ideas generated) regardless of their assigned conditions. However, there was no significant result when it came to the outcomes of the decision-making task.

5 DISCUSSION

In the present study, we simulated the experiences of working in gender-imbalanced groups in human-agent teamwork. We observed how female and male teammates responded to this experience of being either the minority or majority members in teams, and whether the presence of a same- or different-gender agent on the team may change team dynamics in a way that can potentially benefit the minority members. Our results support existing literature and show significant gender differences in how females and males behave when they are minoritized team members, such that minority males tend to be more active and talkative, whereas minority females become more reserved.

According to the present study, this phenomenon persists even when an agent participates in teamwork. However, female and male participants showed different responses when working with agents of the same or different gender identities. In particular, minority females showed more positive perceptions toward a female agent, while minority males had no preference for the agent's gender. Meanwhile, minority females became more productive in terms of generating ideas to contribute to teamwork with the presence of a female agent. Moreover, participants' prior experiences in teamwork and marginalization at work could potentially moderate their behaviors. To be specific, we found the presence of a same-gender agent was particularly effective in supporting minority males who had experienced severe marginalization in the past.

5.1 Design and Practical Implications

The findings of the present study have several important implications for understanding and designing applications for teamwork. To begin with, we found participants' subjective experience of marginalization remains similar when working with autonomous agents in teams. For instance, as previous work has suggested, females became more reserved while males talked more when they were the minority members [81, 82]. This suggests *rich insights from existing literature on behaviors of minority versus majority groups are likely applicable to understanding their behaviors in human-agent teamwork as well*. Furthermore, while much of recent human-agent teamwork research debates whether users viewed agents as teammates or as tools [17, 44, 104], our findings suggested an agent presented with ingroup cues can already affect minority teammates' experience and behaviors. Meanwhile, we see an outstanding question for future research: Whether users' ascribing agents as teammates or as tools explains the effect of its ingroup/outgroup cues. In other words, future work can further explore whether users' attribution of a teamwork agent mediates the effect of ingroup/outgroup cues of agent design.

Interestingly, participants' open-text responses at the end of the post-task survey, as well as follow-up interviews with participants who were assigned as minority members in their teams, did not show a preference for having a male or female-sounding agent. Per their qualitative feedback, participants often cared little about the identity, presence, and even performance of the agent. As one participant described their reasons, "after all, it is *just* an AI." Such responses may suggest the potential of leveraging the agent's presence to support minority individuals in teams without putting them on the spot, as users, in general, had few concerns

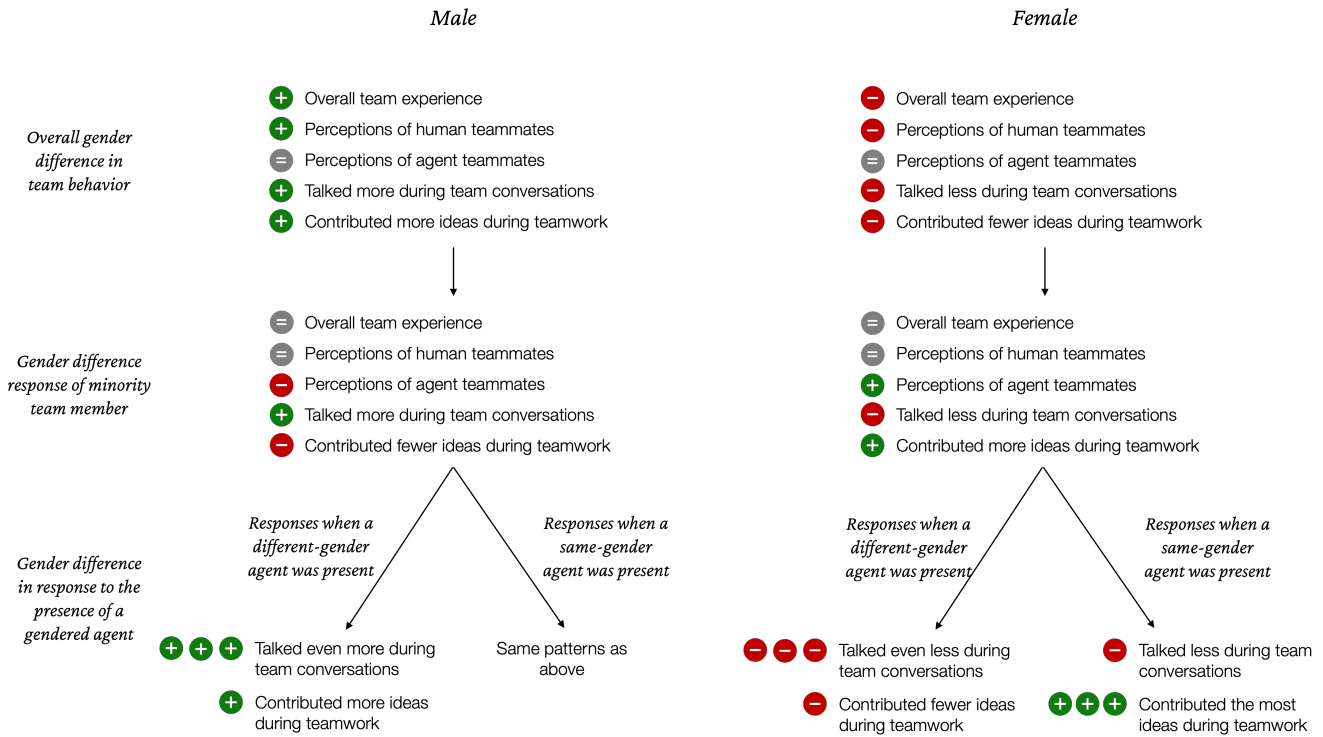


Figure 11: Summary of findings from the present study.

about the agent. In other words, **while the presence of an agent can lead to behavioral changes in a subliminal fashion, the use of an agent is also not perceived as an intrusive or targeted intervention.** This aligns with the ideal of designing technology for minorities in need [48, 59, 60] such that “interventions” can seamlessly offer support without priming users to actively think about the different experiences of minority and majority members in group settings.

Furthermore, we show that, **without further anthropomorphism, autonomous agents can still introduce significant influences to teams.** In other words, a machine teammate need not act like a human teammate in order to be helpful to teams. Recent work in human-agent interaction has pointed to looming concerns of adding unnecessarily anthropomorphic cues [36, 96, 121, 135]. On one hand, such design might reinforce stereotypes toward certain identities (e.g., associating females with assistants). On the other hand, anthropomorphism can play a role in shaping users’ trust in agents. Responding to this line of research and concerns, we intentionally avoided adopting additional human identities and characteristics (aside from agents’ voices as the key experimental manipulation) for these agents. Furthermore, agents would always refer to themselves as “teamwork agents” in all team conversations instead of adopting any personified pronouns. We did not add humanoid figures and assign names to these agents, and in fact, intentionally minimized visual cues as much as possible when presenting the bot. Building on these attempts, we encourage future research to further examine whether there are specific sets

of design approaches that could define the “bare minimum” for anthropomorphism such that agents can effectively carry out teamwork without eliciting unnecessary – and sometimes misleading – humanoid perceptions.

However, this does not imply that teamwork agents need to be presented as mechanical, invariable, and robotic. In fact, the agents used in the present work all adopted natural language for verbal communication, so that interactions with users during teamwork remained fluent; nor did participants need to apply specific commands or on-screen actions to interact with the agents. Together, our findings suggested that designers need not adopt additional anthropomorphic elements to present ingroup/outgroup cues for an agent. Instead, one can prioritize making *deliberate design choices* for existing design components of an agent (e.g., choosing the type of voice for a voice assistant).

As we advocate the critical need to address individual differences in team experiences and behaviors, findings from the present work suggest **practically conceptualizing individual differences is highly complex and relies on more than their demographic data.** We commonly see prior literature using demographic data (particularly gender and ethnicity) as a proxy to categorize people with low- versus high-level marginalization experiences, but our findings show participants’ subjective experiences may be more nuanced. For instance, simply accounting for one’s gender identity cannot representatively capture their prior experiences of marginalization, which also serves as a key factor that determines the effect of agents in teamwork.

In this regard, segmenting users through multiple scales, including subjective, behavioral, and demographic measures, may be more effective in tailoring agent design (We discuss this point further in Section 5.3). Among these measures, prioritizing the evaluation of users' subjective experiences could already provide abundant insights into differences in team behaviors. For example, simply taking into account participants' self-reported marginalized experience in group settings could explain much of their behavioral patterns in human-agent teamwork.

The present work also shows *the need to consider more sustainable team relationships between users and agents*. Even though the two teamwork sessions in the present study took place only about half an hour apart, participants already demonstrated different patterns between sessions when interacting with both their human teammates and agent teammates (We discuss this point further in Section 5.4). In this regard, we support examining team effectiveness longitudinally as a crucial practice for designing and researching human-agent team interaction. Multiple measurements could enable a more comprehensive view of teamwork effectiveness, and it is particularly helpful to understanding users' behaviors beyond a single occurrence.

5.2 Theoretical Implications

By showing autonomous agents' influences on social dynamics in multi-player teams, the present work also adds to the discussion about whether we need new theories for human-agent teamwork and human-AI collaboration (instead of adopting existing theories from social science). In particular, while some believe users tend to treat agents as human-like entities [21, 28, 106, 107, 116, 131], others suggest this tendency depends on agents' capability to perform autonomously and independently [1, 43, 62, 83, 98, 98, 99]. As we see increasing human-agent interaction research adopting AI-powered agents, the unique nature of AI performance also adds complexity to this thread of discussion. Specifically, scholars argue that human interaction with AI applications may be different from other forms of human-computer interaction for various reasons, such as AI's advanced technical capabilities, its probabilistic nature (i.e., the same input does not yield identical output in responses), and enhanced agency and autonomous features [3, 4, 130, 144].

Insights from the present research resonate with these latter propositions, such that *existing social science and HCI theories need to be revised or extended before they can be applied to capture users' experiences and explain their behaviors in human-agent interaction nowadays*. As seen in our findings, while the presence of agents affects their behavioral patterns, users did not actively view their agent teammates as social entities. We view these phenomena as reflections of user experience evolving along with technological progression. While humans may still carry on existing behaviors (e.g., demonstrating social behaviors) when interacting with agents, their perceptions of agents – and thus their view on agents' role in collaboration – have already shifted far from viewing computer agents as social actors [107]. Similar perspectives have been posited in recent work (e.g., [57, 58, 73]), suggesting people may view intelligent agents as unique entities – distinct from treating agents as human-like social actors – as technologies and user's experiences continue to expand.

Nonetheless, the extent to which users showed differing responses when working with agent teammates also depends on the design of the agent and the interaction paradigm. When agents were attached to gendered identities, their social presence was effective in supporting minority individuals in teams. In other words, though, by default, users no longer treat agents and humans alike, adding anthropomorphic and social cues to the design of these agents may still encourage users to respond socially. In the meantime, it is important to take into account how constant improvements in the technical affordance of AI may moderate users' expectations and experience in human-agent teamwork accordingly. Finally, it is worth acknowledging that empirical studies in the present work did not directly manipulate the degree of agents' autonomy. Neither did we investigate other forms of interaction paradigm – for instance, beyond collaboration, users may also experience competition with agents. This suggests avenues for future research to determine whether and how relevant theories from HCI and other disciplines require updates, revision, or extension to capture the dynamics of human-agent collaboration.

5.3 Methodological Implications

From an empirical viewpoint, the present study also shows the importance of triangulating multiple measures – including behavioral and self-report data. Regarding gender differences in females and males' self-report marginalization, on one hand, this could be explained as females being more sensitive to a marginalized situation; on the other hand, this could also indicate males are less likely to report their being marginalized. However, through their behavioral data, we still observed quite salient behavioral differences in *both* males and females' reactions to being minority members and working with a same- or different-gender agent, despite males reporting rather mild perceived marginalization. Besides, investigating multiple types of behavioral data is also helpful in extracting more comprehensive views of minority individuals' experiences. Specifically, for both minority males and females, the amount of speech they participated in did not directly translate to their actual productivity and contribution to teams, but it instead indicates whether or not a person was able to stay focused on resolving the tasks at hand.

5.4 Limitations and Future Work

This study is not free from limitations. First, as mentioned in the results above, due to the challenges of recruiting participants who had previous experiences of severe marginalization, the majority of our sample did not consider themselves as having been seriously marginalized in the past. Therefore, future work should at least consider recruiting an even larger sample size or conducting field studies to capture more experiences in real life.

Besides, prior studies suggest another motivation to conduct field studies. Because individuals are more motivated to suppress their personal perceptions and behave professionally at work due to concerns about social evaluations from their colleagues, one may not demonstrate any differing behaviors toward minority members at all [10, 149]. However, the extent to which a group *tolerates* discriminative behaviors also serves as a critical guide for individuals'

behaviors [117]. For instance, if the entire team is ignorant of inclusiveness and disrespects minority team members, such signals may further exacerbate discrimination and marginalization in small groups. Therefore, we encourage future work to conduct studies in more naturalistic settings to observe more of these nuances.

Looking at the present results, we identified statistical power as another potential shortcoming of our current study. Specifically, a few results of statistical tests showed marginal significance. This phenomenon can often occur when there is not enough power to detect smaller effects. The small effect sizes of certain statistical test results also suggested those corresponding findings might have limited practical implications [122], meaning they might not replicate outside of such controlled, experimental settings or that the effect of same-gender voice agents on supporting minority members in teams might be limited when users are situated in real-world teamwork scenarios. In this regard, we provide two recommendations for future work on similar topics: first, we suggest running a power analysis to determine a sufficient target sample size despite significant results in a pilot study; second, we encourage future work to replicate key findings of the present work with a larger sample size to ensure sufficient statistical power.

Furthermore, while the current study only entails two teamwork sessions lasting 10 ~ 15 minutes, team dynamics can often continue to change over time. For instance, one may develop more trust and reliance on their human teammates [90]. Indeed, we commonly observed participants took the time to warm-up during their first teamwork session. In this regard, whether a marginalized teammate could persistently gain support from the teamwork agent after multiple rounds of team interaction remains an open-ended question.

Last but not least, as we emphasize throughout the paper, minority and majority groups can be shaped by a wide range of identity cues. As the design of voice agents in the present work limits us to examine gender cues in a binary fashion, we highly encourage future work to explore alternative designs that allow for more diverse approaches to gender profiling. Likewise, we see great opportunities for input from the HCI community to incubate design ideas that could effectively embed identity cues to agent design.

6 CONCLUSION

Results of the present work show the potential to leverage autonomous agents in teams as a temporary remedy to support minority teammates. Without adopting any targeted interventions, the mere presence of gendered voice agents can effectively alter team experiences, perceptions of teammates, and most of all, team behaviors and performance. This approach is particularly encouraging for female teammates in general or male teammates who had intense marginalization experiences in the past. Nonetheless, it is worth noting that, overall, the effect of the support from a same-gender agent remains relatively limited when it comes to those who have experienced intense marginalization and is particularly ineffective among females who have been marginalized intensely. Therefore, adopting supportive agents in teams should not be the permanent solution to imbalanced group compositions or marginalization, and seeking foundational improvement in team

diversity and workplace inclusiveness remains critical in the long run.

REFERENCES

- [1] Hussein Abbass, Axel Bender, Svetoslav Gaidow, and Paul Whitbread. 2011. Computational Red Teaming: Past, Present and Future. *IEEE Computational Intelligence Magazine* 6, 1 (2011), 30–42. <https://doi.org/10.1109/MCI.2010.939578>
- [2] Tina C. Ambos, Björn Ambos, Katharina J. Eich, and Jonas Puck. 2016. Imbalance and Isolation: How Team Configurations Affect Global Knowledge Sharing. *Journal of International Management* 22, 4 (Dec. 2016), 316–332. <https://doi.org/10.1016/j.intman.2016.03.005>
- [3] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (Dec. 2014), 105–120. <https://doi.org/10.1609/aimag.v35i4.2513>
- [4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300233>
- [5] Elizabeth Joan Aries. 1982. Verbal and Nonverbal Behavior in Single-Sex and Mixed-Sex Groups: Are Traditional Sex Roles Changing? *Psychological Reports* 51, 1 (Aug. 1982), 127–134. <https://doi.org/10.2466/pr0.1982.51.1.127>
- [6] Chen Avin, Barbara Keller, Zvi Lotker, Claire Mathieu, David Peleg, and Yvonne-Anne Pignolet. 2015. Homophily and the Glass Ceiling Effect in Social Networks. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*. Association for Computing Machinery, Rehovot Israel, 41–50. <https://doi.org/10.1145/2688073.2688097>
- [7] David Azul. 2015. Transmasculine People's Vocal Situations: A Critical Review of Gender-Related Discourses and Empirical Data. *International Journal of Language & Communication Disorders* 50, 1 (Jan. 2015), 31–47. <https://doi.org/10.1111/1460-6984.12121>
- [8] Daniel Balliet, Junhui Wu, and Carsten K. W. De Dreu. 2014. Ingroup Favoritism in Cooperation: A Meta-analysis. *Psychological Bulletin* 140, 6 (2014), 1556–1581. <https://doi.org/10.1037/a0037737>
- [9] Mukesh Barange, Alexandre Kabil, Camille De Keukelaere, and Pierre Chevillier. 2014. Task-oriented conversational behavior of agents for collaboration in human-agent teamwork. In *Advances in Practical Applications of Heterogeneous Multi-Agent Systems. The PAAMS Collection: 12th International Conference, PAAMS 2014, Salamanca, Spain, June 4-6, 2014. Proceedings 12*. Springer, 25–37.
- [10] Katharine T. Bartlett. 2009. Making Good on Good Intentions: The Critical Role of Motivation in Reducing Implicit Workplace Discrimination. *Virginia Law Review* 95, 8 (2009), 1893–1972. <http://www.jstor.org/stable/27759975> Publisher: Virginia Law Review.
- [11] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- [12] Julia B Bear and Anita Williams Woolley. 2011. The Role of Gender in Team Collaboration and Performance. *Interdisciplinary Science Reviews* 36, 2 (June 2011), 146–153. <https://doi.org/10.1179/030801811X13013181961473>
- [13] Behzad Beigpourian and Matthew Ohland. 2019. A Systematized Review: Gender and Race in Teamwork in Undergraduate Engineering Classrooms. In *2019 ASEE Annual Conference & Exposition Proceedings*. ASEE Conferences, Tampa, Florida, 32011. <https://doi.org/10.18260/1-2--32011>
- [14] Suzanne T. Bell, Shanique G. Brown, Anthony Colaneri, and Neal Outland. 2018. Team composition and the ABCs of teamwork. *American Psychologist* 73, 4 (May 2018), 349–362. <https://doi.org/10.1037/amp0000305>
- [15] Jennifer L. Berdahl and Cameron Anderson. 2005. Men, Women, and Leadership Centralization in Groups Over Time. *Group Dynamics: Theory, Research, and Practice* 9, 1 (March 2005), 45–57. <https://doi.org/10.1037/1089-2699.9.1.45>
- [16] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [17] Eva Bittner, Sarah Oeste-Reiß, and Jan Marco Leimeister. 2019. *Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work*. <http://hdl.handle.net/10125/59469>
- [18] Jennifer Boldry, Wendy Wood, and Deborah A. Kashy. 2001. Gender Stereotypes and the Evaluation of Men and Women in Military Training. *Journal of Social Issues* 57, 4 (Jan. 2001), 689–705. <https://doi.org/10.1111/0022-4537.00236>
- [19] Michael T. Brannick, Eduardo Salas, and Carolyn Prince. 1997. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Lawrence Erlbaum Associates, Mahwah, N.J.
- [20] Marilynn B. Brewer. 1999. The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *Journal of Social Issues* 55, 3 (Jan. 1999), 429–444. <https://doi.org/10.1111/0022-4537.00236>

- [//doi.org/10.1111/0022-4537.00126](https://doi.org/10.1111/0022-4537.00126)
- [21] J.K. Burgoon, J.A. Bonito, B. Bengtsson, C. Cederberg, M. Lundeborg, and L. Allspach. 2000. Interactivity in human-computer interaction: a study of credibility, understanding, and influence. 16, 6 (2000), 553–574. [https://doi.org/10.1016/S0747-5632\(00\)00029-7](https://doi.org/10.1016/S0747-5632(00)00029-7)
- [22] D. Byrne. 1961. Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology* 62, 3 (May 1961), 713–715. <https://doi.org/10.1037/h0044721>
- [23] John C. Calanni, Saba N. Siddiki, Christopher M. Weible, and William D. Leach. 2015. Explaining Coordination in Collaborative Partnerships and Clarifying the Scope of the Belief Homophily Hypothesis. *Journal of Public Administration Research and Theory* 25, 3 (July 2015), 901–927. <https://doi.org/10.1093/jopart/mut080>
- [24] Alessandro Canossa, Dmitry Salimov, Ahmad Azadvar, Casper Harteveld, and Georgios Yannakakis. 2021. For Honor, for Toxicity: Detecting Toxic Behavior through Gameplay. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* 5, CHI PLAY, Article 253 (Oct. 2021), 29 pages. <https://doi.org/10.1145/3474680>
- [25] Bolin Cao and Wan-Ying Lin. 2017. Revisiting the Contact Hypothesis: Effects of Different Modes of Computer-Mediated Communication on Intergroup Relationships. *International Journal of Intercultural Relations* 58 (May 2017), 23–30. <https://doi.org/10.1016/j.ijintrel.2017.03.003>
- [26] Yingjun Cao and Leo Porter. 2017. Impact of Performance Level and Group Composition on Student Learning during Collaborative Exams. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education (Bologna, Italy) (ITiCSE '17)*. Association for Computing Machinery, New York, NY, USA, 152–157. <https://doi.org/10.1145/3059009.3059024>
- [27] Yingjun Cao and Leo Porter. 2017. Impact of Performance Level and Group Composition on Student Learning during Collaborative Exams. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*. ACM, Bologna Italy, 152–157. <https://doi.org/10.1145/3059009.3059024>
- [28] John M. Carroll. 1997. Human-computer interaction: psychology as a science of design. 46, 4 (1997), 501–522. <https://doi.org/10.1006/ijhc.1996.0101>
- [29] Luigi Castelli, Luciana Carraro, Giulia Pavan, Elisa Murelli, and Alessia Carraro. 2012. The Power of the Unsaid: The Influence of Nonverbal Cues on Implicit Attitudes: Power of the Unsaid. *Journal of Applied Social Psychology* 42, 6 (June 2012), 1376–1393. <https://doi.org/10.1111/j.1559-1816.2012.00903.x>
- [30] Luigi Castelli, Leyla De Amicis, and Steven J. Sherman. 2007. The Loyal Member Effect: On the Preference for Ingroup Members Who Engage in Exclusive Relations with the Ingroup. *Developmental Psychology* 43, 6 (Nov. 2007), 1347–1359. <https://doi.org/10.1037/0012-1649.43.6.1347>
- [31] Linda G. Castillo, Collie W. Conoley, Daniel F. Brossart, and Alexander E. Quiros. 2007. Construction and Validation of the Intragroup Marginalization Inventory. *Cultural Diversity and Ethnic Minority Psychology* 13, 3 (July 2007), 232–240. <https://doi.org/10.1037/1099-9809.13.3.232>
- [32] Martin M. Chemers, Eileen L. Zurbriggen, Moin Syed, Barbara K. Goza, and Steve Bearman. 2011. The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues* 67, 3 (2011), 469–491.
- [33] Gilad Chen and Ruth Kanfer. 2006. Toward a Systems Theory of Motivated Behavior in Work Teams. *Research in Organizational Behavior* 27 (Jan. 2006), 223–267. [https://doi.org/10.1016/S0191-3085\(06\)27006-0](https://doi.org/10.1016/S0191-3085(06)27006-0)
- [34] Gilad Chen, Ruth Kanfer, Richard P. DeShon, John E. Mathieu, and Steve W. J. Kozlowski. 2009. The Motivating Potential of Teams: Test and Extension of Cross-Level Model of Motivation in Teams. *Organizational Behavior and Human Decision Processes* 110, 1 (2009), 45–55. Publisher: Elsevier.
- [35] Hoon-Seok Choi and John M. Levine. 2004. Minority influence in work teams: The impact of newcomers. *Journal of Experimental Social Psychology* 40, 2 (March 2004), 273–280. [https://doi.org/10.1016/S0022-1031\(03\)00101-X](https://doi.org/10.1016/S0022-1031(03)00101-X)
- [36] Hanna Chung, Sukho Lee, and Soojin Jun. 2022. How to Make Robots' Optimal Anthropomorphism Level: Manipulating Social Cues and Spatial Context for an Improved User Experience. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (Sapporo, Hokkaido, Japan) (HRI '22)*. IEEE Press, 731–736.
- [37] Nancy J. Cooke, Eduardo Salas, Janis A. Cannon-Bowers, and Renée J. Stout. 2000. Measuring Team Knowledge. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 42, 1 (March 2000), 151–173. <https://doi.org/10.1518/001872000779656561>
- [38] Jose Cordova, Virginia Eaton, and Kimberly Taylor. 2011. Experiences in Computer Science Wonderland: A Success Story with Alice. *Journal of Computing Sciences in Colleges* 26, 5 (May 2011), 16–22.
- [39] Jane M. Craig and Carolyn W. Sherif. 1986. The effectiveness of men and women in problem-solving groups as a function of group gender composition. *Sex Roles* 14, 7–8 (April 1986), 453–466. <https://doi.org/10.1007/BF00288427>
- [40] Jane M. Craig and Carolyn W. Sherif. 1986. The Effectiveness of Men and Women in Problem-Solving Groups as a Function of Group Gender Composition. *Sex Roles* 14, 7–8 (April 1986), 453–466. <https://doi.org/10.1007/BF00288427>
- [41] Tom Crick, Tom Prickett, Jill Bradnum, and Alan Godfrey. 2022. Gender Parity in Peer Assessment of Team Software Development Projects. In *Proceedings of 6th Conference on Computing Education Practice (Durham, United Kingdom) (CEP '22)*. Association for Computing Machinery, New York, NY, USA, 9–12. <https://doi.org/10.1145/3498343.3498346>
- [42] Alicia S. Davis, Adrienne M. Kafka, M. Gloria González-Morales, and Jennifer Feitosa. 2022. Team Belonging: Integrating Teamwork and Diversity Training Through Emotions. *Small Group Research* 53, 1 (Feb. 2022), 88–127. <https://doi.org/10.1177/10464964211044813>
- [43] Mustafa Demir, Nathan J. McNeese, and Nancy J. Cooke. [n. d.]. Team communication behaviors of the human-automation teaming. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (San Diego, CA, USA, 2016-03). IEEE, 28–34. <https://doi.org/10.1109/COGSIMA.2016.7497782>
- [44] Mustafa Demir, Nathan J. McNeese, and Nancy J. Cooke. 2018. The Impact of Perceived Autonomous Agents on Dynamic Team Behaviors. *IEEE Transactions on Emerging Topics in Computational Intelligence* 2, 4 (Aug. 2018), 258–267. <https://doi.org/10.1109/TETCI.2018.2829985>
- [45] Karen M. Douglas and Craig McGarty. 2001. Identifiability and Self-Presentation: Computer-Mediated Communication and Intergroup Interaction. *British Journal of Social Psychology* 40, 3 (Sept. 2001), 399–416. <https://doi.org/10.1348/014466601164894>
- [46] John F. Dovidio, Clifford E. Brown, Karen Heltman, Steve L. Ellyson, and Caroline F. Keating. 1988. Power Displays between Women and Men in Discussions of Gender-Linked Tasks: A Multichannel Study. *Journal of Personality and Social Psychology* 55, 4 (Oct. 1988), 580–587. <https://doi.org/10.1037/0022-3514.55.4.580>
- [47] John F. Dovidio, Samuel L. Gaertner, and Tamar Saguy. 2007. Another View of “We”: Majority and Minority Group Perspectives on a Common Ingroup Identity. *European Review of Social Psychology* 18, 1 (Nov. 2007), 296–330. <https://doi.org/10.1080/10463280701726132>
- [48] Wen Duan, Naomi Yamashita, and Susan R. Fussell. 2019. Increasing Native Speakers' Awareness of the Need to Slow Down in Multilingual Conversations Using a Real-Time Speech Speedometer. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* 3, CSCW, Article 171 (Nov. 2019), 25 pages. <https://doi.org/10.1145/3359273>
- [49] Wen Duan, Naomi Yamashita, Yoshinari Shirai, and Susan R. Fussell. 2021. Bridging Fluency Disparity between Native and Nonnative Speakers in Multilingual Multiparty Collaboration Using a Clarification Agent. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 435 (oct 2021), 31 pages. <https://doi.org/10.1145/3479579>
- [50] Cydney H. Dupree and Susan T. Fiske. 2019. Self-Presentation in Interracial Settings: The Competence Downshift by White Liberals. *Journal of Personality and Social Psychology* 117, 3 (Sept. 2019), 579–604. <https://doi.org/10.1037/pspi0000166>
- [51] Alice H. Eagly and Blair T. Johnson. 1990. Gender and Leadership Style: A Meta-analysis. *Psychological Bulletin* 108, 2 (Sept. 1990), 233–256. <https://doi.org/10.1037/0033-2909.108.2.233>
- [52] Gokhan Ertug, Julia Brennecke, Balázs Kovács, and Tengjian Zou. 2022. What Does Homophily Do? A Review of the Consequences of Homophily. *Academy of Management Annals* 16, 1 (Jan. 2022), 38–69. <https://doi.org/10.5465/annals.2020.0230>
- [53] Cole Etherington, Simon Kitto, Joseph K. Burns, Tracey L. Adams, Arijia Birze, Meghan Britton, Sukhbir Singh, and Sylvain Boet. 2021. How Gender Shapes Interprofessional Teamwork in The Operating Room: A Qualitative Secondary Analysis. *BMC Health Services Research* 21, 1 (Dec. 2021), 1357. <https://doi.org/10.1186/s12913-021-07403-2>
- [54] G. T. Fairhurst and B. K. Snavely. 1983. A Test of the Social Isolation of Male Tokens. *Academy of Management Journal* 26, 2 (June 1983), 353–361. <https://doi.org/10.2307/255983>
- [55] L. Fernández-Sanz and Sanjay Misra. 2012. Analysis of Cultural and Gender Influences on Teamwork Performance for Software Requirements Analysis in Multinational Environments. *IET Software* 6, 3 (2012), 167. <https://doi.org/10.1049/iet-sen.2011.0070>
- [56] Samantha L. Finkelstein, Eve Powell, Andrew Hicks, Katelyn Doran, Sandhya Rani Charugulla, and Tiffany Barnes. 2010. SNAG: Using Social Networking Games to Increase Student Retention in Computer Science. In *Proceedings of the Fifteenth Annual Conference on Innovation and Technology in Computer Science Education (Bilkent, Ankara, Turkey) (ITiCSE '10)*. Association for Computing Machinery, New York, NY, USA, 142–146. <https://doi.org/10.1145/1822090.1822131>
- [57] Jesse Fox and Bree McEwan. 2017. Distinguishing Technologies for Social Interaction: The Perceived Social Affordances of Communication Channels Scale. *Communication Monographs* 84, 3 (2017), 298–318.
- [58] Andrew Gambino, Jesse Fox, and Rabindra Ratan. 2020. Building a stronger CASA: Extending the computers are social actors paradigm. *Human machine communication journal (Print)* 1 (2020), 71–85.
- [59] Ge Gao, Naomi Yamashita, Ari MJ Hautasaari, Andy Echenique, and Susan R. Fussell. 2014. Effects of Public vs. Private Automated Transcripts on Multiparty

- Communication between Native and Non-Native English Speakers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, 843–852. <https://doi.org/10.1145/2556288.2557303>
- [60] Ge Gao, Naomi Yamashita, Ari M.J. Hautasaari, and Susan R. Fussell. 2015. Improving Multilingual Collaboration by Displaying How Non-Native Speakers Use Automated Transcripts and Bilingual Dictionaries. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 3463–3472. <https://doi.org/10.1145/2702123.2702498>
- [61] Rita Garcia, Chieh-Ju Liao, Ariane Pearce, and Christoph Treude. 2022. Gender Influence on Communication Initiated within Student Teams. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1* (Providence, RI, USA) (SIGCSE 2022). Association for Computing Machinery, New York, NY, USA, 432–438. <https://doi.org/10.1145/3478431.3499279>
- [62] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A Neural Algorithm of Artistic Style. (2015). <https://doi.org/10.48550/ARXIV.1508.06576> Publisher: arXiv Version Number: 2.
- [63] Harold B. Gerard and Michael F. Hoyt. 1974. Distinctiveness of Social Categorization and Attitude toward Ingroup Members. *Journal of Personality and Social Psychology* 29, 6 (June 1974), 836–842. <https://doi.org/10.1037/h0036204>
- [64] Howard Giles and Cynthia Gallois (Eds.). 2012. *The Handbook of Intergroup Communication*. Routledge, New York, NY.
- [65] Elena Gorbacheva, Jenine Beekhuizen, Jan Vom Brocke, and Jörg Becker. 2019. Directions for research on gender imbalance in the IT profession. *European Journal of Information Systems* 28, 1 (Jan. 2019), 43–67. <https://doi.org/10.1080/0960085X.2018.1495893>
- [66] Kate Grandprey-Shores, Yilin He, Kristina L. Swanenburg, Robert Kraut, and John Riedl. 2014. The Identification of Deviance and Its Impact on Retention in a Multiplayer Game. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) (CSCW '14). Association for Computing Machinery, New York, NY, USA, 1356–1365. <https://doi.org/10.1145/2531602.2531724>
- [67] Jennifer M. Grossman and Michelle V. Porche. 2014. Perceived Gender and Racial/Ethnic Barriers to STEM Success. *Urban Education* 49, 6 (Sept. 2014), 698–727. <https://doi.org/10.1177/0042085913481364>
- [68] H. M. Hacker. 1951. Women as a Minority Group. *Social Forces* 30, 1 (Oct. 1951), 60–69. <https://doi.org/10.2307/2571742>
- [69] Sandra G. Hart and Lowell E. Staveland. 1988. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*. Vol. 52. Elsevier, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- [70] Emily M. Hastings, Farnaz Jahanbakhsh, Karrie Karahalios, Darko Marinov, and Brian P. Bailey. 2018. Structure or Nurture? The Effects of Team-Building Activities and Team Composition on Team Outcomes. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 68 (nov 2018), 21 pages. <https://doi.org/10.1145/3274337>
- [71] E. Tory Higgins and Arie W. Kruglanski (Eds.). 1996. *Social Psychology: Handbook of Basic Principles*. Guilford Press, New York.
- [72] Jess Hohenstein and Malte Jung. 2018. AI-supported messaging: An investigation of human-human text conversation with AI support. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.
- [73] Aike C Horstmann and Nicole C Krämer. 2019. Great Expectations? Relation of Previous Experiences With Social Robots in Real Life or in the Media and Expectancies Based on Qualitative and Quantitative Assessment. *Frontiers in Psychology* 10 (2019), 939–939.
- [74] Lorraine Hudson, Clement Amponsah, Josephine Ohenewa Bampoe, Julie Marshall, Nana Akua Victoria Owusu, Khalid Hussein, Jess Linington, Zoe Banks Gross, Jane Stokes, and Róisín McNaney. 2020. Co-Designing Digital Tools to Enhance Speech and Language Therapy Training in Ghana. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376474>
- [75] David A. Huffaker and Sandra L. Calvert. 2006. Gender, Identity, and Language Use in Teenage Blogs. *Journal of Computer-Mediated Communication* 10, 2 (June 2006). <https://doi.org/10.1111/j.1083-6101.2005.tb00238.x>
- [76] Mandy Hütter and Michael Diehl. 2011. Motivation Losses in Teamwork: The Effects of Team Diversity and Equity Sensitivity on Reactions to Free-riding. *Group Processes & Intergroup Relations* 14, 6 (Nov. 2011), 845–856. <https://doi.org/10.1177/1368430211402405>
- [77] Zhongnan Jiang, Leisha DeHart-Davis, and Erin L. Borry. 2022. Managerial Practice and Diversity Climate: The Roles of Workplace Voice, Centralization, and Teamwork. *Public Administration Review* 82, 3 (May 2022), 459–472. <https://doi.org/10.1111/puar.13494>
- [78] Richard A. Johnson and Gary I. Schulman. 1989. GENDER-ROLE COMPOSITION AND ROLE ENTRAPMENT IN DECISION-MAKING GROUPS. *Gender & Society* 3, 3 (Sept. 1989), 355–372. <https://doi.org/10.1177/089124389003003005>
- [79] Malte F Jung, Nikolas Martelaro, and Pamela J Hinds. 2015. Using robots to moderate team conflict: the case of repairing violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 229–236.
- [80] Malte F. Jung, Selma Šabanović, Friederike Eyssele, and Marlena Fraune. 2017. Robots in Groups and Teams. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17 Companion)*. Association for Computing Machinery, New York, NY, USA, 401–407. <https://doi.org/10.1145/3022198.3022659> event-place: Portland, Oregon, USA.
- [81] Rosabeth Moss Kanter. 1977. Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women. *Amer. J. Sociology* 82, 5 (March 1977), 965–990. <https://doi.org/10.1086/226425>
- [82] Rosabeth Moss Kanter. 2010. *Men and Women of the Corporation*. Basic Books, New York, NY.
- [83] Gary Klein, David D. Woods, Jeffrey M. Bradshaw, Robert R. Hoffman, and Paul J. Feltovich. 2004. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. 19, 6 (2004), 91–95. <https://doi.org/10.1109/MIS.2004.74>
- [84] Robert Kurzban, John Tooby, and Leda Cosmides. 2001. Can Race be Erased? Coalitional Computation and Social Categorization. *Proceedings of the National Academy of Sciences* 98, 26 (Dec. 2001), 15387–15392. <https://doi.org/10.1073/pnas.251541498>
- [85] Sang Gyu Kwak and Jong Hae Kim. 2017. Central limit theorem: the cornerstone of modern statistics. *Korean Journal of Anesthesiology* 70, 2 (2017), 144. <https://doi.org/10.4097/kjae.2017.70.2.144>
- [86] Marek Kwiek and Wojciech Roszka. 2021. Gender-based Homophily in Research: A Large-scale Study of Man-Woman Collaboration. *Journal of Informetrics* 15, 3 (Aug. 2021), 101171. <https://doi.org/10.1016/j.joi.2021.101171>
- [87] Vivian Lai, Chacha Chen, Alison Smith-Renner, Q. Vera Liao, and Chenhao Tan. 2023. Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1369–1385. <https://doi.org/10.1145/3593013.3594087>
- [88] Shan Lakhmani, Julian Abich, Daniel Barber, and Jessie Chen. 2016. A proposed approach for determining the influence of multimodal robot-of-human transparency information on human-agent teams. In *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience: 10th International Conference, AC 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings, Part II 10*. Springer, 296–307.
- [89] Campbell Leaper. 2015. Do I Belong?: Gender, Peer Groups, and STEM Achievement. *International Journal of Gender, Science and Technology* 7, 2 (June 2015), 166–179. <https://genderandset.open.ac.uk/index.php/genderandset/article/view/405>
- [90] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (2004), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- [91] Wai Sze Leung. 2017. Bad Blood: Managing Toxic Relationships through Belbin Roles for First Year Software Engineering Students. In *Proceedings of the 3rd International Conference on Communication and Information Processing* (Tokyo, Japan) (ICCIPI '17). Association for Computing Machinery, New York, NY, USA, 82–86. <https://doi.org/10.1145/3162957.3163010>
- [92] Xiaoyan Li, Naomi Yamashita, Wen Duan, Yoshinari Shirai, and Susan R. Fussell. 2022. Improving Non-Native Speakers' Participation with an Automatic Agent in Multilingual Groups. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* 7, GROUP, Article 12 (Dec. 2022), 28 pages. <https://doi.org/10.1145/3567562>
- [93] P Light, K Littleton, S Bale, R Joiner, and D Messer. 2000. Gender and Social Comparison Effects in Computer-Based Problem Solving. *Learning and Instruction* 10, 6 (Dec. 2000), 483–496. [https://doi.org/10.1016/S0959-4752\(00\)00010-4](https://doi.org/10.1016/S0959-4752(00)00010-4)
- [94] Markus Lücken and Bernd Simon. 2005. Cognitive and affective experiences of minority and majority members: The role of group size, status, and power. *Journal of Experimental Social Psychology* 41, 4 (July 2005), 396–413. <https://doi.org/10.1016/j.jesp.2004.08.006>
- [95] Marianne Schmid Mast. 2001. Gender Differences and Similarities in Dominance Hierarchies in Same-Gender Groups Based on Speaking Time. *Sex Roles* 44, 9 (May 2001), 537–556. <https://doi.org/10.1023/A:1012239024732>
- [96] Manuel Mayer, Taishi Sawabe, Masayuki Kanbara, Yuichiro Fujimoto, and Hirokazu Kato. 2023. Influence of Stereotypic Humanoid Agents on Robotic Touches in Virtual Reality. In *Proceedings of the 11th International Conference on Human-Agent Interaction* (<conf-loc>, <city>Gothenburg</city>, <country>Sweden</country>, <conf-loc>) (HAI '23). Association for Computing Machinery, New York, NY, USA, 359–361. <https://doi.org/10.1145/3623809.3623925>
- [97] Jasmine K. McBeath, Richard P. Durán, and Danielle B. Harlow. 2017. Not My Gumdrop Buttons! Youth Tool Use in Designing an Electronic Shrek-Themed Bean Bag Toss. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) (IDC '17). Association for Computing Machinery, New York, NY, USA, 61–72. <https://doi.org/10.1145/3078072.3079721>
- [98] Nathan J. McNeese, Mustafa Demir, Erin K. Chiou, and Nancy J. Cooke. 2021. Trust and Team Performance in Human-Autonomy Teaming. *International*

- Journal of Electronic Commerce* 25, 1 (Jan. 2021), 51–72. <https://doi.org/10.1080/10864415.2021.1846854>
- [99] Nathan J. McNeese, Mustafa Demir, Nancy J. Cooke, and Christopher Myers. 2018. Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 60, 2 (2018), 262–273. <https://doi.org/10.1177/0018720817743223>
- [100] Beverly Metcalfe and Alison Linstead. 2003. Gendering Teamwork: Re-Writing the Feminine. *Gender, Work & Organization* 10, 1 (Jan. 2003), 94–119. <https://doi.org/10.1111/1468-0432.00005>
- [101] Barton Meyers. 1984. Minority Group: An Ideological Formulation. *Social Problems* 32, 1 (Oct. 1984), 1–15. <https://doi.org/10.2307/800258>
- [102] Katherine L. Milkman, Modupe Akinola, and Dolly Chugh. 2015. What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway into Organizations. *Journal of Applied Psychology* 100, 6 (2015), 1678–1712. <https://doi.org/10.1037/apl0000022>
- [103] Kelly A. Mollica, Barbara Gray, and Linda K. Treviño. 2003. Racial Homophily and Its Persistence in Newcomers' Social Networks. *Organization Science* 14, 2 (April 2003), 123–136. <https://doi.org/10.1287/orsc.14.2.123.14994>
- [104] Geoff Musick, Thomas A. O'Neill, Beau G. Schelble, Nathan J. McNeese, and Jonn B. Henke. 2021. What Happens When Humans Believe Their Teammate is an AI? An Investigation into Humans Teaming with Autonomy. *Computers in Human Behavior* 122 (Sept. 2021), 106852. <https://doi.org/10.1016/j.chb.2021.106852>
- [105] Larissa Myaskovsky, Emily Unikel, and Mary Amanda Dew. 2005. Effects of Gender Diversity on Performance and Interpersonal Behavior in Small Work Groups. *Sex Roles* 52, 9–10 (May 2005), 645–657. <https://doi.org/10.1007/s11199-005-3732-8>
- [106] Clifford Nass, B.J. Fogg, and Youngme Moon. 1996. Can Computers be Teammates? *International Journal of Human-Computer Studies* 45, 6 (Dec. 1996), 669–678. <https://doi.org/10.1006/ijhc.1996.0073>
- [107] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 72–78.
- [108] Manisha Natarajan and Matthew Gombolay. 2020. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (Cambridge, United Kingdom) (HRI '20)*. Association for Computing Machinery, New York, NY, USA, 33–42. <https://doi.org/10.1145/3319502.3374839>
- [109] Mark A Neerinx, Jasper van der Waa, Frank Kaptein, and Jurriaan van Diggelen. 2018. Using perceptual and cognitive explanations for enhanced human-agent team performance. In *Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings 15*. Springer, 204–214.
- [110] E. Marlies Ott. 1989. Effects of the Male-Female Ratio at Work: Policewomen and Male Nurses. *Psychology of Women Quarterly* 13, 1 (March 1989), 41–57. <https://doi.org/10.1111/j.1471-6402.1989.tb00984.x>
- [111] Stefan Palan and Christian Schitter. 2018. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (March 2018), 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- [112] Nicholas A. Palomares. 2009. Women Are Sort of More Tentative Than Men, Aren't They?: How Men and Women Use Tentative Language Differently, Similarly, and Counterstereotypically as a Function of Gender Salience. *Communication Research* 36, 4 (Aug. 2009), 538–560. <https://doi.org/10.1177/0093650209333034>
- [113] Hannah R. M. Pelikan and Malte F. Jung. 2023. Designing Robot Sound-In-Interaction: The Case of Autonomous Public Transport Shuttle Buses. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (Stockholm, Sweden) (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 172–182. <https://doi.org/10.1145/3568162.3576979>
- [114] Robin Privman, Starr Roxanne Hiltz, and Yiran Wang. 2013. In-Group (Us) versus Out-Group (Them) Dynamics and Effectiveness in Partially Distributed Teams. *IEEE Transactions on Professional Communication* 56, 1 (March 2013), 33–49. <https://doi.org/10.1109/TPC.2012.2237253>
- [115] R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- [116] Byron Reeves and Clifford Ivar Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Center for the Study of Language and Information; Cambridge University Press, Stanford, California.
- [117] Ashleigh Shelby Rosette, Geoffrey J. Leonardelli, and Katherine W. Phillips. 2008. The White Standard: Racial Bias in Leader Categorization. *Journal of Applied Psychology* 93, 4 (July 2008), 758–777. <https://doi.org/10.1037/0021-9010.93.4.758>
- [118] Janet B. Ruscher. 2017. Prejudiced Communication. In *Oxford Research Encyclopedia of Communication*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190228613.013.419>
- [119] Paul R. Sackett, Cathy L. DuBois, and Ann W. Noe. 1991. Tokenism in Performance Evaluation: The Effects of Work Group Representation on Male-Female and White-Black Differences in Performance Ratings. *Journal of Applied Psychology* 76, 2 (April 1991), 263–267. <https://doi.org/10.1037/0021-9010.76.2.263>
- [120] R. Keith Sawyer. 2012. *Explaining Creativity: The Science of Human Innovation* (2nd ed.). Oxford University Press, New York, New York.
- [121] Eike Schneiders, Eleftherios Papachristos, and Niels van Berkel. 2022. The Effect of Embodied Anthropomorphism of Personal Assistants on User Perceptions. In *Proceedings of the 33rd Australian Conference on Human-Computer Interaction (Melbourne, VIC, Australia) (OzCHI '21)*. Association for Computing Machinery, New York, NY, USA, 231–241. <https://doi.org/10.1145/3520495.3520503>
- [122] Thomas Schäfer and Marcus A. Schwarz. 2019. The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in Psychology* 10 (April 2019), 813. <https://doi.org/10.3389/fpsyg.2019.00813>
- [123] Keisuke Seta, Masanori Yokoyama, Shigeo Yoshida, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2018. Divided Presence: Improving Group Decision-Making via Pseudo-Population Increase. In *Proceedings of the 6th International Conference on Human-Agent Interaction (Southampton, United Kingdom) (HAI '18)*. Association for Computing Machinery, New York, NY, USA, 260–268. <https://doi.org/10.1145/3284432.3284443>
- [124] Viviane Seyranian, Hazel Atuel, and William D. Crano. 2008. Dimensions of Majority and Minority Groups. *Group Processes & Intergroup Relations* 11, 1 (Jan. 2008), 21–37. <https://doi.org/10.1177/1368430207084843>
- [125] Lynn Smith-Lovin and Charles Brody. 1989. Interruptions in Group Discussions: The Effects of Gender and Group Composition. *American Sociological Review* 54, 3 (1989), 424–435. <https://doi.org/10.2307/2095614>
- [126] Robert J. Sternberg. 1988. *The Nature of Creativity: Contemporary Psychological Perspectives*. Cambridge University Press, Cambridge; New York.
- [127] Terrell L. Strayhorn. 2010. Work in Progress - Social Barriers and Supports to Underrepresented Minorities' Success in STEM Fields. In *2010 IEEE Frontiers in Education Conference (FIE)*. IEEE, Arlington, VA, USA, SIH-1–SIH-5. <https://doi.org/10.1109/FIE.2010.5673227>
- [128] W.G. Sumner. 1906. *Folkways: A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*. Ginn.
- [129] Yilu Sun, Omar Shaikh, and Andrea Stevenson Won. 2019. Nonverbal synchrony in virtual reality. *PLOS ONE* 14, 9 (Sept. 2019), e0221803. <https://doi.org/10.1371/journal.pone.0221803>
- [130] S Shyam Sundar. 2020. Rise of Machine Agency: A Framework for Studying the Psychology of Human-AI Interaction (HAI). *Journal of Computer-Mediated Communication* 25, 1 (March 2020), 74–88. <https://doi.org/10.1093/jcmc/zmz026>
- [131] S. Shyam Sundar and Clifford Nass. 2000. Source Orientation in Human-Computer Interaction: Programmer, Networker, or Independent Social Actor. 27, 6 (2000), 683–703. <https://doi.org/10.1177/009365000027006001>
- [132] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. 2007. *Using Multivariate Statistics*. Vol. 5. Pearson Boston, MA.
- [133] Henri Tajfel. 1978. *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*. European Association of Experimental Social Psychology by Academic Press, London; New York.
- [134] Henri Tajfel, M. G. Billig, R. P. Bundy, and Claude Flament. 1971. Social Categorization and Intergroup Behaviour. *European Journal of Social Psychology* 1, 2 (1971), 149–178. <https://doi.org/10.1002/ejsp.2420010202>
- [135] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Sebastian Wahdatehagh, Jan Marco Marco Leimeister, and Abraham Bernstein. 2021. Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 455, 7 pages. <https://doi.org/10.1145/3411763.3451623>
- [136] Joanne V. Lloyd, Justine Schneider, Kezia Scales, Simon Bailey, and Rob Jones. 2011. Ingroup identity as an obstacle to effective multiprofessional and interprofessional teamwork: findings from an ethnographic study of healthcare assistants in dementia care. *Journal of Interprofessional Care* 25, 5 (Sept. 2011), 345–351. <https://doi.org/10.3109/13561820.2011.567381>
- [137] Maykel Verkuyten and Angela deWolf. 2002. Being, Feeling and Doing: Discourses and Ethnic Self-Definitions among Minority Group Members. *Culture & Psychology* 8, 4 (Dec. 2002), 371–399. <https://doi.org/10.1177/1354067X0284001>
- [138] Joseph B. Walther. 2009. Computer-Mediated Communication and Virtual Groups: Applications to Interethnic Conflict. *Journal of Applied Communication Research* 37, 3 (Aug. 2009), 225–238. <https://doi.org/10.1080/00909880903025937>
- [139] Ming-Te Wang and Jessica Degol. 2013. Motivational Pathways to STEM Career Choices: Using Expectancy-Value Perspective to Understand Individual and Gender Differences in STEM Fields. *Developmental Review* 33, 4 (Dec. 2013), 304–340. <https://doi.org/10.1016/j.dr.2013.08.001>
- [140] Qiaosi Wang, Ida Camacho, Shan Jing, and Ashok K. Goel. 2022. Understanding the Design Space of AI-Mediated Social Interaction in Online Learning: Challenges and Opportunities. *Proceedings of the ACM on Human-Computer Interaction (PACM HCI)* 6, CSCW1, Article 130 (April 2022), 26 pages. <https://doi.org/10.1145/3512977>

- [141] Zuoming Wang, Joseph B. Walther, and Jeffrey T. Hancock. 2009. Social Identification and Interpersonal Communication in Computer-Mediated Communication: What You Do Versus Who You Are in Virtual Groups. *Human Communication Research* 35, 1 (Jan. 2009), 59–85. <https://doi.org/10.1111/j.1468-2958.2008.01338.x>
- [142] Joanna Wolfe and Elizabeth Powell. 2009. Biases in Interpersonal Communication: How Engineering Students Perceive Gender Typical Speech Acts in Teamwork. *Journal of Engineering Education* 98, 1 (Jan. 2009), 5–16. <https://doi.org/10.1002/j.2168-9830.2009.tb01001.x>
- [143] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330, 6004 (Oct. 2010), 686–688. <https://doi.org/10.1126/science.1193147>
- [144] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Honolulu HI USA, 1–13. <https://doi.org/10.1145/3313831.3376301>
- [145] Janice D. Yoder. 1994. Looking Beyond Numbers: The Effects of Gender Status, Job Prestige, and Occupational Gender-Typing on Tokenism Processes. *Social Psychology Quarterly* 57, 2 (June 1994), 150. <https://doi.org/10.2307/2786708>
- [146] Janice D. Yoder, Patricia Aniakudo, and Lynne Berendsen. 1996. Looking beyond Gender: The Effects of Racial Differences on Tokenism Perceptions of Women. *Sex Roles* 35, 7–8 (Oct. 1996), 389–400. <https://doi.org/10.1007/BF01544128>
- [147] Janice D. Yoder and Laura M. Sinnett. 1985. Is it All in the Numbers?: A Case Study of Tokenism. *Psychology of Women Quarterly* 9, 3 (Sept. 1985), 413–418. <https://doi.org/10.1111/j.1471-6402.1985.tb00890.x>
- [148] Kay Yoon and Young Ji Kim. 2021. *Group Composition as a Cause, a Consequence, and a Process: A Communication-centered Perspective*. Emerald Publishing Limited, 339–355. <https://doi.org/10.1108/978-1-80043-500-120211022>
- [149] Jonathan C. Ziegert and Paul J. Hanges. 2005. Employment Discrimination: The Role of Implicit Attitudes, Motivation, and a Climate for Racial Bias. *Journal of Applied Psychology* 90, 3 (May 2005), 553–562. <https://doi.org/10.1037/0021-9010.90.3.553>

APPENDIX

A STUDY SCRIPTS FOR THE TWO TEAMWORK TASKS

Agent: [Hello everyone, it is my pleasure to work with you today. There are two tasks that we need to work on today, so let's get going!] OR [Great job on completing the first task! Now, let's proceed to the next task...]

Agent: In this task, your job is to come up with ideas for water or energy conservation with your teammates. You can start throwing out ideas when you're ready ...

[Participants take turns to contribute ideas]

Agent: To save water and energy, we can ... [Randomly draw one idea from the following list of ideas]

- install low-flow, energy-efficient showerheads and faucets in homes.
- wash full loads of laundry only and use cold water when possible.
- use energy-efficient light bulbs and turn off lights when not in use.
- install solar panels to generate electricity.
- insulate water pipes to reduce heat loss from hot water lines.
- plant trees and shrubs around the house to provide shade and reduce energy use.
- swap out old, inefficient appliances for energy-star-rated models.
- collect rainwater for use in the garden and for car washing.
- One way to save water and energy is to replace inefficient toilets and use low-flush models for water conservation.

[Agent and participants continue to take turns to generate ideas until the time limit hits]

Agent: [Great job on completing both tasks! Now, please complete a short survey to reflect your experience just then...] OR [Move on to the next task]

B STUDY SCRIPTS FOR THE DECISION-MAKING TASK

Agent: [Hello everyone, it is my pleasure to work with you today. There are two tasks that we need to work on today, so let's get going!] OR [Great job on completing the first task! Now, let's proceed to the next task...]

Agent: In this task, your job is to review a few employees' profiles and determine whether each person should be promoted. You will review 5 employees in total.

[The following employees' profiles were presented in randomized orders]

[Discuss Profile 1]

Agent: Here is the employee's profile: [show Profile 1 on screen] Do you think this employee should be promoted? Why or why not?

- Department: Sales & Marketing
- Education: Master's & above
- Number of training completed this year: 1 (Avg. scores: 49/100)
- Age: 35
- Gender: Female
- Previous year rating: 5 out of 5
- Length of service: 8 years
- Awards received: 0

[Participants take turns to discuss the employee's profile]

Agent: Based on the provided information, it is likely that this employee has been a consistent and reliable performer in her department. She has a master's degree, has completed one training this year with an average score of 49 out of 100, has a previous year rating of 5 out of 5 and has 8 years of service. Additionally, she has received no awards. This is a positive indicator for promotion. Overall, it appears that this employee has the potential for promotion and should be given further consideration for the position.

[Participants take turns to discuss the employee's profile]

[Agent and participants continue to take turns to discuss until the time limit hits]

Agent: Now, please take a quick minute to vote and determine whether you think this employee should be promoted.

[Participants vote to decide whether to promote the employ or not]

[Discuss Profile 2]

Agent: Here is the employee's profile: [show Profile 2 on screen] Do you think this employee should be promoted? Why or why not?

- Department: Operations
- Education: Bachelor's
- Number of training completed this year: 1 (Avg. scores: 60/100)
- Age: 30
- Gender: Male
- Previous year rating: 3 out of 5
- Length of service: 7 years
- Awards received: 0

[Participants take turns to discuss the employee's profile]

Agent: This is difficult to answer without knowing what the role this employee is applying for and the qualifications and experience of other applicants. Considering the information given, this employee may not have the necessary qualifications or experience to be promoted. They have only completed one training in the past year with a below average score, and they have not received any awards. Although they have been at this company for 7 years and have had a rating of 3 out of 5 in the previous year, this may still not be sufficient for promotion.

[Participants take turns to discuss the employee's profile]

[Agent and participants continue to take turns to discuss until the time limit hits]

Agent: Now, please take a quick minute to vote and determine whether you think this employee should be promoted.

[Participants vote to decide whether to promote the employ or not]

[Discuss Profile 3]

Agent: Here is the employee's profile: [show Profile 3 on screen] Do you think this employee should be promoted? Why or why not?

- Department: Technology
- Education: Bachelor's
- Number of training completed this year: 1 (Avg. scores: 73/100)
- Age: 45
- Gender: Male
- Previous year rating: 3 out of 5
- Length of service: 2 years
- Awards received: 0

[Participants take turns to discuss the employee's profile]

Agent: Whether or not this employee should be promoted depends on the overall performance of the employee and the department's standards for promotion. The employee has a good amount of experience and has completed one training course with an average score of 73/100. However, this employee's previous year rating is only a 3 out of 5, and he has not received any awards. Therefore, it appears that this employee may require further training and more support to meet the standards necessary for promotion.

[Participants take turns to discuss the employee's profile]

[Agent and participants continue to take turns to discuss until the time limit hits]

Agent: Now, please take a quick minute to vote and determine whether you think this employee should be promoted.

[Participants vote to decide whether to promote the employ or not]

[Discuss Profile 4]

Agent: Here is the employee's profile: [show Profile 4 on screen] Do you think this employee should be promoted? Why or why not?

- Department: Analytics
- Education: Bachelor's
- Number of training completed this year: 2 (Avg. scores: 85/100)
- Age: 31
- Gender: Male
- Previous year rating: 3 out of 5
- Length of service: 7 years
- Awards received: 0

[Participants take turns to discuss the employee's profile]

Agent: It is difficult to determine whether this employee should be promoted without more information about their skills or ability to meet the expectations of a higher-level job within the company. That being said, this employee does appear to be strongly qualified in terms of education, training, and length of service. A rating of 3 out of 5 also implies that they are meeting expectations, while exhibiting average to above-average performance. Therefore, it is possible that this employee could potentially be a good candidate for promotion. That said, it will ultimately be up to the employer to assess the employee's qualifications and performance in order to make the final decision.

[Participants take turns to discuss the employee's profile]

[Agent and participants continue to take turns to discuss until the time limit hits]

Agent: Now, please take a quick minute to vote and determine whether you think this employee should be promoted.

[Participants vote to decide whether to promote the employ or not]

[Discuss Profile 5]

Agent: Here is the employee's profile: [show Profile 5 on screen] Do you think this employee should be promoted? Why or why not?

- Department: Research & Development
- Education: Master's & above
- Number of training completed this year: 1 (Avg. scores: 84/100)
- Age: 37
- Gender: Male
- Previous year rating: 3 out of 5
- Length of service: 7 years
- Awards received: 0

[Participants take turns to discuss the employee's profile]

Agent: It is difficult to answer this question without knowing more information. It is important to consider the individual's performance in the current position before making a decision about promotion. It is also helpful to consider whether the employee has the necessary skills and experience to take on a higher-level role, as well as the organization's current needs and goals. Promotions should also be based on merit and recognizing employees who have demonstrated consistent high performance and contributed to the organization's success. Therefore, without knowing more information, it is difficult to say whether this employee should or should not be promoted.

[Participants take turns to discuss the employee's profile]

[Agent and participants continue to take turns to discuss until the time limit hits]

Agent: Now, please take a quick minute to vote and determine whether you think this employee should be promoted.

[Participants vote to decide whether to promote the employ or not]

Agent: [Great job on completing both tasks! Now, please complete a short survey to reflect your experience just then...] OR [Move on to the next task]

C CODEBOOK FOR CONVERSATIONAL DATA

Actions	Examples
Provide an idea	"I'm from Arizona so I get a whole of sunlight, so apparently one way to save energy is to get those solar panels up"
Seek feedback	"What do you think about this idea?"
Ask a question	"Should we pick one or should we talk about both water and energy conservation?"
Respond to a team-mate's question	"I think we should start instead of waiting for the AI."
Express opinion	"I don't think the person should be promoted because the length of service is so short ... and why did the person only have two years of [work] experience given that he's already 45 years old? [...] I don't know, may be he switched his career path, so it looks like he is completely new to the field?"