Int. J. Human–Computer Studies 181 (2024) 103146

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Exploring how gender-anonymous voice avatars influence women's performance in online computing group work

Dominic Kao [a,*], Syed T. Mubarrat [a], Amogh Joshi [a], Swati Pandita [b], Christos Mousas [a], Hai-Ning Liang [c], Rabindra Ratan [d]

[a] Purdue University, USA
[b] California Institute of Technology, USA
[c] Xi'an Jiaotong-Liverpool University, China
[d] Michigan State University, USA

ARTICLE INFO

Keywords:
Avatar
Audio
Voice
Stereotype threat
Group work
Computing

ABSTRACT

We investigate how gender-anonymous voice avatars influence women's performance in online computing group work. Female participants worked with two male confederates. Voices were filtered according to four voice gender anonymity conditions: (1) All unmasked, (2) Male confederates masked, (3) Female participant masked, and (4) All masked. When only male confederates used masked voices (compared to all unmasked), female participants spoke for a longer period of time and scored higher on computing problems. When everyone used masked voices (compared to all unmasked), female participants spoke for a longer period of time, spoke more words, and scored higher on computing problems. Effects were not significant on subjective measures and one behavioral measure. We discuss the implications for virtual interactions between people.

1. Introduction

Teams working together remotely became commonplace during the COVID-19 pandemic (Brynjolfsson et al., 2020), driven by factors such as cost reduction and flexibility in work-life balance (Ferreira et al., 2021), with many companies now offering positions that are permanently remote (Smith, 2022). Online education is also growing—the global market of $269.87 billion USD in 2021 is expected to increase to $585.48 billion USD in 2027 (Renub Research, 2022). Communication platforms such as Zoom, Microsoft Teams, Cisco Webex, Slack, and Discord are now common in businesses, education, and entertainment (Blagojević, 2022). However, communication on such platforms may reinforce social inequities in the workplace and online education. For example, multiple studies have found that women experience significantly higher zoom fatigue than men (Fauville et al., 2021; Ratan et al., 2022). The present research is concerned with the potential for online teamwork to augment the already high level of disparity for women in computing fields, where fewer than 20% of computer science students are women (Sax et al., 2017; Higher Education Statistics Agency, 2018; Joint Council for Qualifications, 2018), in part due to historical (Henn, 2014) and cultural biases (Abbate, 2012). Further, stereotypes that women do not belong in technological fields engender hostile and sexist environments that hinder equitable participation (Dean, 2007; Cohoon and Aspray, 2006) and trigger stereotype threat, the phenomenon that reminders of a negative stereotype about a social group increase the risk that members of the group will conform to the stereotype or otherwise perform poorly (Koch et al., 2008; Master et al., 2016; Steele, 2010). Stereotype threat—one of the most widely studied topics in social psychology (Schmader et al., 2008) across various domains (Master et al., 2016; Pennington et al., 2016; Walton and Spencer, 2009)—has been implicated in long-standing racial and gender inequalities in academic performance (Osborne, 2001; Goldsmith, 2004; Ellison and Swanson, 2010), but has not been studied extensively in the context of online teamwork. Hence, the present research addresses this gap, recognizing that gender diversity creates better performing teams (Woolley et al., 2010; Gratton et al., 2007; Turner, 2009) and offering a theoretically grounded approach to mitigating stereotype threat and promoting more equitable online participation for women in technological team meetings.

In particular, we build on research suggesting that avatars—defined as mediated representations of humans used to facilitate real-time interactions (Nowak and Fox, 2018)—play an important role in stereotype threat effects within mediated communication. For example, one study found that African-American participants using race-anonymized

* Corresponding author.
E-mail addresses: kaod@purdue.edu (D. Kao), smubarra@purdue.edu (S.T. Mubarrat), joshi134@purdue.edu (A. Joshi), sp2333@cornell.edu (S. Pandita), cmousas@purdue.edu (C. Mousas), HaiNing.Liang@xjtlu.edu.cn (H.-N. Liang), rar@msu.edu (R. Ratan).

https://doi.org/10.1016/j.ijhcs.2023.103146
Received 15 March 2023; Received in revised form 15 August 2023; Accepted 7 September 2023
Available online 15 September 2023
1071-5819/

avatars persisted significantly longer in a cognitive task with other ostensible participants than those using race-revealing avatars (Lee, 2009). Another study found that both men and women participants who used a female avatar in a threat condition (i.e., in a group with two ostensible male avatars) performed worse on a math task than participants who used a male avatar (Lee et al., 2014). Another study found that using a female avatar in VR harmed math task performance and math confidence, while using a male avatar buffered against these stereotype threat effects (Peck et al., 2020). In all of these studies, a visual avatar was assigned to the participants, and although a few studies have examined the role of avatar customization in these effects (Fordham et al., 2020; Ratan and Sah, 2015), no studies of which we are aware have examined the role of non-visual avatars in stereotype threat effects. Addressing this gap, the present research focuses on mediated representations of user voices, which can be considered *voice avatars*, and which carry important cues of user identity that are potentially stereotyped, such as gender.

Although voice avatars are a novel technology in the context of online teamwork, there is significant evidence that voice-based communication in online gaming environments triggers hostility and sexism toward women (Kuznekoff and Rose, 2013; McLean and Griffiths, 2019). Further, although participants in an online meeting can use a visual avatar on many platforms, few if any offer the option for voice avatars, and the effects of using such avatars are largely unknown. Hence, the lack of research on voice avatars and stereotype threat represents a significant gap.

The technology required to mask or modify user voices is advancing rapidly (Kao et al., 2021), so we may expect that team communication platforms will offer these tools in the future, but the best approaches to mitigating stereotyping and stereotype threat through such tools are unclear. As a first step toward exploring this technology and the role of stereotype threat in online team communication, we focus on a simple comparison of gender-revealing and gender-masking voice avatars.[1] Hence, for the purposes of this study, we created a faux online meeting platform that allowed our self-identifying female participants to use either a gender-revealing or gender-masking voice avatar to interact with two ostensible male teammates (actually recorded confederates). Results suggest that gender-masking voice avatars buffered the effects of stereotype threat as reflected by the amount participants spoke and performance on computing problems, though subjective measures were not found to be affected. Together, this large-scale preregistered study suggests that stereotype threat is likely rampant in online technical meetings and that voice (and visual) avatars can likely be used to promote gender equity in such teamwork across a variety of contexts (virtual classes, business meetings, and entertainment).

2. Related work

2.1. Stereotype threat

Stereotype threat is a type of psychological discomfort triggered when people with a negatively-stereotyped social identity feel at risk of confirming the negative stereotype and consequently experience adverse effects in the stereotype-relevant domain (Steele, 1997; Steele and Aronson, 1995). In the seminal study on stereotype threat Steele and Aronson (1995), African-American students underperformed on the verbal Graduate Records Examination compared to their White counterparts when the test was framed as an assessment of intellectual ability or students were asked to record their race/ethnicity in a questionnaire prior to completing the task. However, no racial performance differences were found when the test was framed as a problem-solving exercise or the race/ethnicity question was administered after the test. This performance difference likely occurred because African-Americans are often negatively stereotyped as inferior to White students in intellectual domains, and thus the test framing that increased the negative stereotype's salience induced a psychological threat that inhibited their performance. These findings provide evidence that stereotypical views and beliefs about a group can negatively influence the intellectual functioning of individual group members.

2.1.1. Domains, mechanisms, and triggers of stereotype threat

As one of the most widely studied topics in social psychology (Schmader et al., 2008), stereotype threat has been shown to hinder performance in a wide variety of domains, including mathematics, computer science, athletics (Birney et al., 2020; Ellison and Swanson, 2010; Goldsmith, 2004; Osborne, 2001; Pennington et al., 2016; Schmader, 2012; Steele and Aronson, 1995; Stone et al., 2012; Thoman et al., 2013; Walton and Spencer, 2009), intelligence tests (Croizet and Claire, 1998), engineering (Bell et al., 2003; Logel et al., 2009), mathematical reasoning (Inzlicht and Ben-Zeev, 2000; McGlone and Aronson, 2006; Schmader, 2002; Spencer et al., 1999), political knowledge (McGlone et al., 2006), memory (Hess and Hinson, 2006; Levy, 1996), and English (Birney et al., 2020). But in addition to directly harming performance (Shapiro and Neuberg, 2007), stereotype threat affects a variety of other factors, such as sense of belonging (Cheryan and Plaut, 2010; Murphy et al., 2007; Smith et al., 2013; Thoman et al., 2013; Walton and Cohen, 2007), motivation (Cheryan et al., 2009; Murphy et al., 2007; Steele et al., 2002a), interest (Smith et al., 2007), and career intentions (Cheryan et al., 2011; Davies et al., 2005; Gupta and Bhawe, 2007). And although many stereotype threat studies have focused on racial and gender stereotypes, the effect has been found for numerous stigmatized groups, such as Latinos (Gonzales et al., 2002), Native Americans (Osborne, 2001), older adults (Hess and Hinson, 2006; Levy, 1996), and people with low socioeconomic status (Croizet and Claire, 1998).

There are numerous (debated) psychological mechanisms for stereotype threat Shapiro and Neuberg (2007), such as anxiety, stress, and arousal (Birney et al., 2020; Blascovich et al., 2001; Osborne, 2007; Palomares, 2008; Steele, 2010), which may hinder attention (Hasher and Zacks, 1988) and cognitive efficiency (Eysenck and Calvo, 1992). Studies have found evidence of the mediating role of anxiety on stereotype threat outcomes as reflected in measures of communication fluency (Palomares, 2009; McGlone and Pfiester, 2015) and psychophysiological activity—such as mean arterial pressure reactivity (Blascovich et al., 2001), skin conductance, surface skin temperature, and diastolic blood pressure (Osborne, 2007). Many situational factors can trigger stereotype threat (Hoyt and Murphy, 2016; Murphy and Taylor, 2012; Steele et al., 2002b), such as the test's diagnosticity—the extent to which a test is framed as an evaluation of a negatively stereotyped individual's skills/abilities (Bell et al., 2003; Chalabaev et al., 2008; Croizet and Claire, 1998; Gonzales et al., 2002; Johns et al., 2008; McGlone et al., 2006; McGlone and Pfiester, 2015; Steele and Aronson, 1995; Taylor and Walton, 2011)—or if a test/exam is perceived as highly important (Danaher and Crandall, 2008; Kray et al., 2002; Schmader et al., 2009). Stereotype threat results from even subtle cues to test diagnosticity, such as telling participants they will receive performance feedback after a test (Martens et al., 2006; Schmader and Johns, 2003; Steele and Aronson, 1995) or that their strengths and weaknesses are being investigated (Johns et al., 2008; Steele and Aronson, 1995).

Furthermore, the type of diagnosticity makes a difference—e.g., women suffered more from stereotype threat when they were led to believe their leadership ability (stereotypically masculine skill) was

[1] We recognize that gender is often treated as a non-binary concept (Bem, 1981; Ansara and Berger, 2016; Lips, 2017). This paper focuses on stereotype threat related to women in computing contexts, using men as the comparison gender category, consistent with previous gender-stereotype research (e.g., Hargittai and Shafer (2006), Nguyen and Ryan (2008), Giammarco et al. (2015)). This approach does not intend to dichotomize gender, but leaves room for future research to examine stereotype threat and other gender categorizations.

being evaluated compared to being told relationship maintenance ability (stereotypically feminine skill) was being evaluated (McGlone and Pfiester, 2015). Stereotype threat is also triggered when a stereotyped identity is made salient (Beilock et al., 2006, 2007; Blascovich et al., 2001; Brown and Pinel, 2003; Delgado and Prieto, 2008; Johns et al., 2005; Keller, 2002; McGlone and Pfiester, 2015; McIntyre et al., 2003; O'Brien and Crandall, 2003; Osborne, 2007; Rosenthal et al., 2007; Smith et al., 2007; Spencer et al., 1999; Stone and McWhinnie, 2008). For instance, women who were told that men are better than women in mathematical domains exhibited stereotype threat effects during a subsequent computer science programming task (Smith et al., 2007) and math problems (Beilock et al., 2007; Keller, 2002). Implicit cues to stereotype relevance can also trigger stereotype threat, for example, by telling female participants that their performance on a math test would be compared to men (Delgado and Prieto, 2008; Rosenthal et al., 2007) or that the objective of an experiment was to investigate differences between men and women in athletics (Stone and McWhinnie, 2008) and mathematics (Brown and Pinel, 2003; Johns et al., 2005; McIntyre et al., 2003; Spencer et al., 1999). Similarly, job descriptions utilizing more masculine-themed words compared to feminine or neutral words predicted lower belonging for potential female candidates (Stout and Dasgupta, 2011) and caused women to report lower scores of expected belongingness and job appeal (Gaucher et al., 2011). And perhaps even more implicitly, being asked to indicate gender, race/ethnicity, or age in questionnaires (Danaher and Crandall, 2008; McGlone and Aronson, 2006; Schmader and Johns, 2003; Shih et al., 1999; Steele and Aronson, 1995), being in physical (Cheryan et al., 2009) or virtual (Cheryan et al., 2011) environments decorated with stereotypically masculine objects, or simply being numerically underrepresented (e.g., women taking a test in a majority-male room) increases the likelihood of stereotype threat effects (Beaton et al., 2007; Inzlicht and Ben-Zeev, 2000, 2003; Inzlicht and Good, 2006; Johns et al., 2008; Murphy et al., 2007; Sekaquaptewa and Thompson, 2003).

Relatedly, the presence or behavior of others—such as someone who makes sexist remarks (Logel et al., 2009) or simply a male experimenter (McGlone et al., 2006)—can also trigger stereotype threat (Maass et al., 2008; Palomares, 2009; Stone and McWhinnie, 2008). Similarly, women were more likely to experience stereotype threat after being led to believe they were competing against a man in chess (Maass et al., 2008) or a shooter video game (Fordham et al., 2020). Because people possess multiple social identities (e.g., gender, age, race/ethnicity, socioeconomic status, etc.), stereotype threat also occurs when a social identity relevant to the stereotyped context is made more salient than other social identities (Rosenthal and Crisp, 2006). For example, Asian women suffered from stereotype threat during a math test after filling out a questionnaire about their gender identity, but not when they answered a questionnaire about their ethnic identity (Shih et al., 1999). Similarly, female students primed to be more self-aware about their gender identity performed worse on a standardized spatial reasoning test compared to women who were made aware of their identity as students (McGlone and Aronson, 2006).

2.2. Mitigating stereotype threat online

Given the overwhelming evidence that stereotype threat occurs widely in society, especially among students, researchers should focus on developing approaches to combat it. Small effective steps to mitigate stereotype threat improves overall well-being, reduces worries about being devalued based on group membership (e.g., ethnicity, gender), facilitates meaningful engagement with peers, instructors, and classroom learning activities, and consequently reduces achievement gaps by 30%–40% (Cohen et al., 2012) in the short and long term (Cohen et al., 2006, 2009). Approaches to mitigating stereotype threat generally require altering the psychological environment of the individual (Steele et al., 2002b). For example, just as framing a test as diagnostic triggers stereotype threat, framing a test as generic will counteract it (Steele and Aronson, 1995). In contrast, the present research focuses on stereotype threat mitigation through social-identity mechanisms. Namely, just as increasing a stereotyped social identity's salience triggers stereotype threat, reducing cues to social identity mitigates it (Garcia and Cohen, 2013). The present research focuses on a context of performance where such social identity cues are particularly poignant and also malleable: online group collaborations.

Group interactions and collaborations are increasingly taking place online across many contexts, from undergraduate education (Parker and Lenhart, 2011) to academic research conferences (Wu et al., 2022) to work meetings in general (Karl et al., 2022). In all of these contexts, participants create and share content and engage with others socially through digital self-representations that often reflect their social identities (Boyd and Ellison, 2007; Pegg et al., 2018). Similar to social identity offline, online social identity can be threatened by social cues in the digital environment, and so approaches to mitigate stereotype threat offline can also work in online contexts (Chang et al., 2019). However, digital environments are different because individuals are afforded greater control over how other users perceive them by selectively presenting (or hiding) aspects of their personal identity as a part of their online identity—something that is more difficult to achieve in offline, face-to-face communication (Walther, 1996). Hence, we focus on how digital self-representations online may be used to mitigate stereotype threat.

2.2.1. Avatars influence stereotype threat

In online communication environments, avatars are the fundamental vehicle of users' social identities. Drawing from a well-respected definition that applies broadly across media modalities (Nowak and Fox, 2018), we define avatars here as mediated representations of humans used to interact with others, objects, or environments in real time. Avatars are *mediated*—not the users themselves, but representations of the users—and hence afford a potential for anonymity. In other words, even if an avatar appears (or sounds) very lifelike and realistic, it may not represent the users' identity characteristics accurately, so inferences drawn about the user from the avatar's depiction may be inaccurate. However, despite this rationale, people still treat avatars' artificially anthropomorphic characteristics as representative of actual human characteristics (Kao, 2019) and stereotype them accordingly (Kaye et al., 2018; Ratan and Sah, 2015). Such misattribution of human traits to non-human entities likely occurs because people tend to obliviously respond to social cues—including those exhibited by machines—following the social norms developed through human–human interaction (Nass and Moon, 2000; Gambino et al., 2020), a phenomenon referred to as the media equation (Reeves and Nass, 1996). Further, when people use avatars, they tend to adopt the avatars' identity characteristics into their self-perception and then behaviorally conform to associated stereotypes (Yee and Bailenson, 2007; Ratan et al., 2020). This phenomenon, referred to as the Proteus effect, helps explain why people are also susceptible to stereotype threat induced with respect to their avatars' identity characteristics. For example, one study found that both men and women participants who used a female avatar in a threat condition (i.e., in a group with two ostensible male avatars) performed worse on a math task than participants who used a male avatar (Lee et al., 2014). Another study found that using a female avatar in VR harmed math task performance and math confidence, while using a male avatar buffered against these stereotype threat effects (Peck et al., 2020). And another study with women participants found that those who customized and used a female avatar (compared to a male avatar) in a sword-fighting Wii game performed worse on a competitive math task after the game, but only if perceived avatar embodiment was low (Ratan and Sah, 2015). These previous studies of avatars and stereotype threat focused on visual avatar characteristics (i.e., the avatar's appearance on the screen). No studies of which we are aware have examined the role of non-visual avatars in stereotype threat effects. This is a major gap in the research because voices—and voice avatars—are a highly stereotyped aspect of social identity in online interactions.

2.3. Stereotyping voice (avatars)

Voice has a unique capacity to reflect social identity, acting as an "auditory face" that others evaluate (Belin et al., 2011) to ascertain the speaker's gender (Strand, 1999; Trent, 1995), age (Sebastian and Bouchard Ryan, 2018), ethnicity (Trent, 1995), personality (McAleer et al., 2014), and even social status (Ko et al., 2006; Sebastian and Bouchard Ryan, 2018). Hence, voice can trigger "linguistic profiling" (Gray, 2012), racism, and sexism from others (Chan and Gray, 2020; McLean and Griffiths, 2019). Unlike with visual self-representations (i.e., avatars), most popular online communication platforms—including Zoom, Microsoft Teams, Cisco Webex, and Discord (Blagojević, 2022)—do not presently offer native tools for users to modify or customize the representation of their voices (i.e., voice avatars). Instead, such platforms simply detect audial information from the microphone input and replicate it through speaker outputs with little or no computational processing that affect how identity cues are presented (Byeon et al., 2022; Johns et al., 2008). Such systems hinder privacy and agency in voice communication, which helps explain why women and minority groups are often reluctant to use their voices online to avoid negative attention and delegitimization of their abilities (Cote, 2017; Kuznekoff and Rose, 2013; McLean and Griffiths, 2019, 2013). This is especially problematic because voice is growing in prominence as a medium of communication not only in online groups, but also in human–computer interaction, such as when communicating with virtual assistants and intelligent agents (Cherif and Lemoine, 2017; Clark et al., 2019; Seaborn et al., 2022; Divekar et al., 2019a,b; Xu et al., 2021; Xu and Warschauer, 2020; Zierau et al., 2020), dictating emails (Shah et al., 2021), navigating websites (Anon, 2020), and online learning (Khan et al., 2022; Miyazoe and Anderson, 2011; Paule-Ruiz et al., 2013).

Voice avatars present a novel opportunity to address this issue by allowing users to decouple identity characteristics presented to others through voices from users' offline identity characteristics. Although some researchers have examined voice avatars or similar concepts related to voice modification or customization in computational settings (e.g., Kao et al. (2022, 2021), Okano et al. (2022)), research on voice avatars in online group communication is practically non-existent. As an early exploratory step in this direction, we focus on masking gender identity through voice avatars in a way that facilitates anonymity for the users, meaning their personal identity cannot be readily ascertained through the information shared online (Wallace, 1999, 2008). Studies suggest that such anonymity can mitigate stereotype threat. For example, masking status cues during a choice-dilemma task led to more equitable group discussion between high and low status participants (Dubrovsky et al., 1991). Similarly, masking gender cues during group discussions among executives led to more equitable decision making between men and women participants (Sproull et al., 1991). Research on visual avatars suggests the same pattern. One study found that African-American participants using race-anonymized avatars persisted significantly longer in a cognitive group task than those using race-revealing avatars (Lee, 2009). Another study found that the African-American participants performed worse on a cognitive task when competing (instead of cooperating) with race-revealing avatars (Lee and Nass, 2012).

2.4. Hypotheses

Our hypotheses are based on an assumed context of online interaction in which an individual—the potential target of stereotype threat—interacts with group members potentially with voice avatars designed to anonymize social identity. In order to test our general expectation that voice avatars designed to anonymize users should mitigate stereotype threat, we pose a series of hypotheses (all included in our preregistration) about the mechanisms and the stereotype threat of performance outcomes. Regarding mechanisms, we draw from Self-Determination Theory (SDT) (Ryan and Deci, 2020) to hypothesize that stereotype threat will have a negative effect on competence (e.g., from worse performance (Steele, 2010)), autonomy (e.g., as a correlate of lower motivation (Deci and Ryan, 1987; Harter, 1981)), and relatedness (e.g., from a lower sense of belonging (Thoman et al., 2013)). Just as racial anonymization has been found to mitigate stereotype threat effects (Lee, 2009), we hypothesize that gender anonymization may similarly positively affect competence, autonomy, and relatedness. Regarding outcomes, we focus on proportion of speaking contribution, an outcome studied in previous research on group dynamics and stereotypes (Dubrovsky et al., 1991; Sproull et al., 1991), as well as on facets related to success on the group task.

Additionally, studies have shown that the perception of being in the minority amplifies stereotype threat effects, e.g., Lee and Nass (2012). In this context, when all group members are not using gender anonymized voices, this should reinforce identity salience and the perception of being in the minority, augmenting stereotype threat effects. Therefore, we hypothesize that there will be an interaction effect such that when all group members are not using gender anonymized voices, stereotype threat effects will be greater than the summed stereotype threat effects of individual group subsets not using gender anonymized voices.

Competence

H1.1: Participant voice masking will lead to higher competence.

H1.2: Groupmate voice masking will lead to higher competence.

H1.3: Interaction effect: Participant unmasking will see a greater reduction in competence with respect to participant masking when faux participants are unmasked compared to when faux participants are masked.

Autonomy

H2.1: Participant voice masking will lead to higher autonomy.

H2.2: Groupmate voice masking will lead to higher autonomy.

H2.3: Interaction effect: Participant unmasking will see a greater reduction in autonomy with respect to participant masking when faux participants are unmasked compared to when faux participants are masked.

Relatedness

H3.1: Participant voice masking will lead to higher relatedness.

H3.2: Groupmate voice masking will lead to higher relatedness.

H3.3: Interaction effect: Participant unmasking will see a greater reduction in relatedness with respect to participant masking when faux participants are unmasked compared to when faux participants are masked.

More generally, we hypothesize that women will be less susceptible to stereotype threat effects when communicating through gender-anonymous voices in computing group work.[2] Therefore, we hypothesize that gender-anonymous voices will increase duration of speaking (stereotype threat causes anxiety (Steele, 2010) which can hinder communication fluency (McGlone and Pfiester, 2015) and decrease speech duration (Wörtwein et al., 2015; Goberman et al., 2011; Laukka et al., 2008)),[3] speed of response (stereotype threat leads to communication tentativeness (McGlone and Pfiester, 2015; Palomares, 2008)), correctness of answers (stereotype threat reduces performance (Steele,

[2] This assumes a gender-anonymous visual representation across all conditions, which was used in this study. This is further discussed in Limitations.

[3] Note that we also test for actual number of words spoken in addition to duration. However, this analysis was *not* contained in our preregistration and hence is not specified with formal hypotheses. This analysis was instead performed as a follow-up to test whether results with respect to speech duration would translate to the actual number of words spoken. See Section 6.3.4 for additional rationale and methodology.

2010)), and respect from others (stereotype threat reduces sense of belonging (Murphy et al., 2007; Thoman et al., 2013)). Because stereotype threat effects result from a complex process involving multiple mechanisms (Schmader et al., 2008), many affective, cognitive, and motivational factors have been hypothesized to mediate stereotype threat effects (Pennington et al., 2016). We hypothesize that competence, autonomy, and relatedness will mediate stereotype threat effects because these factors are predictors of variables known to be influenced by stereotype threat (e.g., competence and autonomy are theorized to be positive predictors of motivation (McAuley et al., 1989), relatedness influences sense of belonging (Mendoza-Denton and Page-Gould, 2008)) (Thoman et al., 2013). Moreover, competence, autonomy, and relatedness are theorized to be universal needs that motivate behavior and may hence affect our outcomes (Ryan and Deci, 2000). (See Fig. 3.)

Duration of Speaking
H4.1: Participant voice masking will lead to higher duration of speaking.
H4.2: Groupmate voice masking will lead to higher duration of speaking.
H4.3: Interaction effect: Participant unmasking will see a greater reduction in duration of speaking with respect to participant masking when faux participants are unmasked compared to when faux participants are masked.
H4.4: Competence will mediate H4.1.
H4.5: Competence will mediate H4.2.
H4.6: Autonomy will mediate H4.1.
H4.7: Autonomy will mediate H4.2.
H4.8: Relatedness will mediate H4.1.
H4.9: Relatedness will mediate H4.2.

Speed in Responding
H5.1: Participant voice masking will lead to higher speed in responding.
H5.2: Groupmate voice masking will lead to higher speed in responding.
H5.3: Interaction effect: Participant unmasking will see a greater reduction in speed in responding with respect to participant masking when faux participants are unmasked compared to when faux participants are masked.
H5.4: Competence will mediate H5.1.
H5.5: Competence will mediate H5.2.
H5.6: Autonomy will mediate H5.1.
H5.7: Autonomy will mediate H5.2.
H5.8: Relatedness will mediate H5.1.
H5.9: Relatedness will mediate H5.2.

Correctness of Responses
H6.1: Participant voice masking will lead to higher correctness of responses.
H6.2: Groupmate voice masking will lead to higher correctness of responses.
H6.3: Interaction effect: Participant unmasking will see a greater reduction in correctness of responses with respect to participant masking when faux participants are unmasked compared to when faux participants are masked.
H6.4: Competence will mediate H6.1.
H6.5: Competence will mediate H6.2.
H6.6: Autonomy will mediate H6.1.
H6.7: Autonomy will mediate H6.2.
H6.8: Relatedness will mediate H6.1.
H6.9: Relatedness will mediate H6.2.

Stereotype Threat Scores
H7.1: Participant voice masking will lead to lower stereotype threat scores.
H7.2: Groupmate voice masking will lead to lower stereotype threat scores.
H7.3: Interaction effect: Participant unmasking will see a greater increase in stereotype threat scores with respect to participant masking when faux participants are unmasked compared to when faux participants are masked.
H7.4: Competence will mediate H7.1.
H7.5: Competence will mediate H7.2.
H7.6: Autonomy will mediate H7.1.
H7.7: Autonomy will mediate H7.2.
H7.8: Relatedness will mediate H7.1.
H7.9: Relatedness will mediate H7.2.

Autonomous Respect Scores
H8.1: Participant voice masking will lead to higher autonomous respect scores.
H8.2: Groupmate voice masking will lead to higher autonomous respect scores.
H8.3: Interaction effect: Participant unmasking will see a greater reduction in autonomous respect scores with respect to participant masking when faux participants are unmasked compared to when faux participants are masked.
H8.4: Competence will mediate H8.1.
H8.5: Competence will mediate H8.2.
H8.6: Autonomy will mediate H8.1.
H8.7: Autonomy will mediate H8.2.
H8.8: Relatedness will mediate H8.1.
H8.9: Relatedness will mediate H8.2.

3. Voice avatar creation software

We wanted to develop voice avatar creation software. To do so, as a first step, we sought to leverage voice-changing software that could facilitate the creation of a gender-anonymous voice. Therefore, we (1) reviewed existing voice-changing software; (2) determined that none of the existing voice-changing software were suitable; and (3) developed and validated our own custom voice changer. Details of this process can be found in Supplementary Materials.

4. Online meeting platform

We developed an online meeting platform compatible in any modern browser. We built our own platform because we wanted fine-grained control over the application—e.g., embedding our voice avatar creation software and recording meeting analytics. See Figs. 1 and 2. Details of the platform and the development process can be found in Supplementary Materials.

5. Methods

5.1. Study preregistration

Our study was preregistered on the Open Science Framework (OSF). Hypotheses, experiment design, data collection, sample size, measures, and analyses are contained in our preregistration.

5.2. Conditions

The study used a 2 × 2 factorial design. Participants worked in groups of 3 (participant + 2 male confederates). All participants self-identified as female. Participants were led to believe that the two other group members were real participants. In reality, the two other group members were prerecorded male confederates. Participants were not told explicitly what the gender of the other group members was. We manipulated participant voice (gender-unmasked vs. gender-masked) and the two group members' voices (gender-unmasked vs. gender-masked). Voice anonymity conditions were as follows:

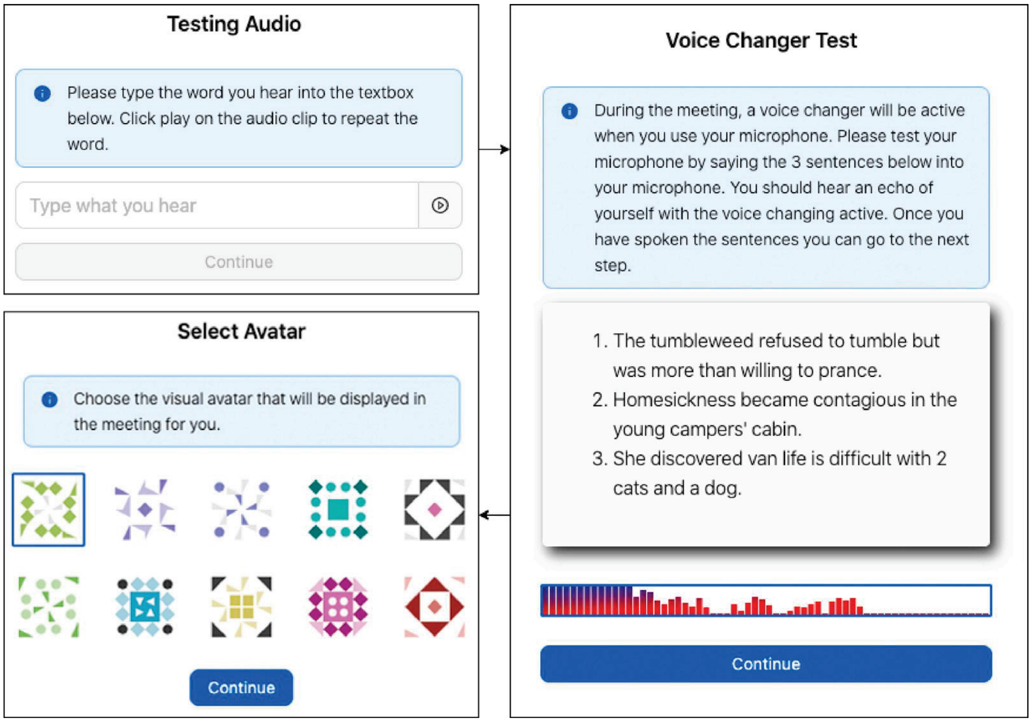

Fig. 1. Meeting setup phase.

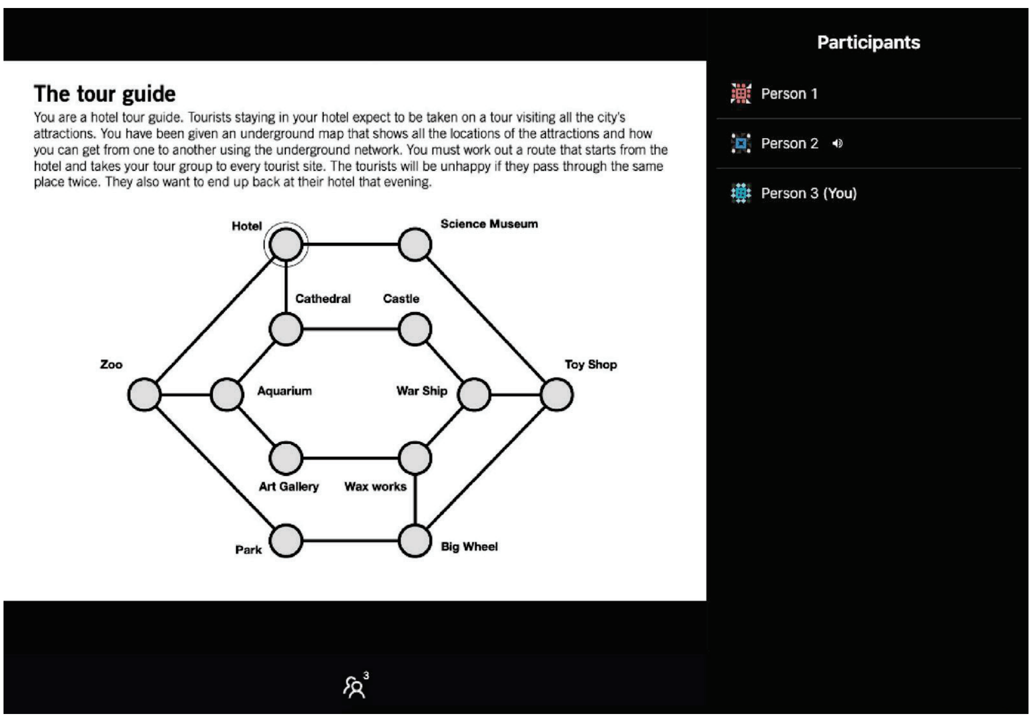

Fig. 2. Meeting.

- **All unmasked**: None of the group members were gender anonymous.
- **Male confederates masked**: Only the two male group members were gender anonymous.
- **Female participant masked**: Only the female participant was gender anonymous.
- **All masked**: All group members were gender anonymous.

Additional details on the specific manner in which stereotype threat was induced, how the male confederate voices were created, and the visual avatar used by meeting participants can be found in Supplementary Materials.

5.3. Measures

Full details and justification for each measure can be found in Supplementary Materials.

5.3.1. Competence

We measured competence using the perceived competence subscale of the Intrinsic Motivation Inventory (IMI) (McAuley et al., 1989).

5.3.2. Autonomy

We measured autonomy using the perceived choice subscale of the IMI (McAuley et al., 1989).

5.3.3. Relatedness

We measured relatedness using the relatedness subscale of the IMI (McAuley et al., 1989).

5.3.4. Stereotype threat measure: Duration of speaking

Duration of speaking was measured by manually inspecting the duration of speech in each participant voice recording, omitting blank silence at the beginning and end.

5.3.5. Stereotype threat measure: Speed in responding

Speed in responding was automatically recorded by the meeting platform. This was the duration of time between when the participant had the ability to begin recording their answer and when the participant actually began recording their answer.

5.3.6. Stereotype threat measure: Correctness of responses

Correctness for each computing problem was determined through audio recordings of participants' responses. Details on how correctness was assessed can be found in Supplementary Materials.

5.3.7. Stereotype threat measure: Stereotype threat survey

We used the "negative stereotype concerns" survey from a previous study on stereotypes in computing (Master et al., 2016).

5.3.8. Stereotype threat measure: Respect from others

We measured "respect from others" (Bartel et al., 2012; Tyler and Blader, 2002). Additional details can be found in Supplementary Materials.

5.4. Sample size determination

We calculated a priori sample size. For 2 × 2 ANOVAs, we used G*Power 3.1 to conduct a power analysis using an effect size of small (0.1), $\alpha = 0.05$, and 95% power. This power analysis found that a sample size of N = 1302 would be necessary.

For the mediation analyses, we use Monte Carlo Power Analysis (Schoemann et al., 2017). We used a model with three parallel mediators, 95% power, 1000 replications, 20,000 Monte Carlo draws per rep, and a 95% confidence level. Correlations between variables are set to 0.2 based on the literature available to us, e.g., a validation of the IMI found moderate correlations between subscales (McAuley et al., 1989). The power analysis found that a sample size of N = 808 would be necessary.

We take the upper bound across both power analyses of N = 1302.

5.5. Participants

1362 participants were recruited from Prolific,[4] with 17 participants removed during data screening (see data screening in Supplementary Materials for rationale). Participants were paid $11 USD per hour. There was no limitation on geographic location. Recruitment details, data screening, participants' demographics, and participants' prior computing experience can be found in Supplementary Materials.

[4] We intentionally recruited a slightly larger number of participants than required (N = 1302) in case of data being removed during screening.

5.6. Design

We used a between-subjects factorial design. There were four possible conditions, and each participant was randomly assigned to one condition. Across conditions, there were approximately equal numbers of participants (M = 336.3, SD = 9.3).

5.7. Procedure

5.7.1. Overview

Here, we provide a brief overview of the procedure. Additional details including verbatim participant instructions can be found in Supplementary Materials. Participants first opened the online meeting link in their browser. After an audio test and microphone test, participants chose their visual avatar (see Section 5.2). After instructions, the participant was paired with two other "participants" (male confederates). Participants were told they would act as judges, while the other two participants were brainstormers.

During the meeting, the group members were presented each of 5 computing problems one at a time. We took the problems from the "CS4FN Puzzle Book"[5] Specifically, we chose puzzles #7, #10, #13, #15, and #17. Each puzzle was presented one at a time and shown on the screen (see Fig. 2). See Section Brainstormer Dialogue Snippet and Example Participant Solutions for the specific dialogue for each problem. Each problem was discussed by the two brainstormers, after which the participant was prompted to record an audio clip of their answer. During recording, participants could hear the output from the voice avatar creation software after a delay in order to reinforce the manipulation.

Afterward, participants filled out a post-survey and were debriefed. Full details of the procedure and justification for study choices can be found in Supplementary Materials.

5.7.2. Brainstormer dialogue snippet and example participant solutions

A complete listing of brainstormer dialogue for all problems and example participant solutions can be found in Supplementary Materials.

5.8. Analysis

We used SPSS 23 and the PROCESS macro (Hayes, 2022) for data analysis. Factorial 2 × 2 ANOVAs were used to study the effects of participant masking and group member masking on IMI perceived competence (H1), IMI perceived choice (H2), and IMI relatedness (H3).[6] We then performed a parallel mediation analysis with condition as a categorical variable using indicator coding (X),[7] competence (M1), autonomy (M2), relatedness (M3), and duration of speaking (Y) (H4). PROCESS model 4 was used (Hayes, 2022). The parallel mediation was repeated using the different outcomes of interest (Y): speed in responding (H5); correctness of responses (H6); stereotype threat scores (H7); and autonomous respect (H8). Interaction effects for H4–H8 were tested separately using 2 × 2 ANOVAs.[8] We used an α of 0.05. Analyses were preregistered. Outlier testing is described in Supplementary Materials.

[5] Problems: https://cs4fndownloads.files.wordpress.com/2016/02/cs4fnpuzzlebook11.pdf Solutions: https://cs4fndownloads.files.wordpress.com/2016/02/cs4fnpuzzles1solutions.pdf

[6] ANOVAs are considered robust to non-normality, especially at larger sample sizes (Blanca et al., 2017).

[7] This means that the first condition ("All unmasked") will be used as the reference category against which all other conditions will be compared.

[8] As specified in our preregistration, these ANOVAs test only for interaction effects and not main effects. The rationale for this is that main effects are already being tested in our mediation analysis, however interaction effects are not.

6. Results

6.1. Manipulation check

The manipulation check consisted of six questions. All manipulation check questions validated the effectiveness of our manipulation. Detailed results of the manipulation check can be found in Supplementary Materials.

Table 1
Descriptives for IMI competence (H1), IMI choice (H2), and IMI relatedness (H3).

	IMI Competence		IMI Choice		IMI Relatedness	
	M	*SD*	*M*	*SD*	*M*	*SD*
Participant Unmasked						
Group Members Unmasked	4.59	1.43	5.94	1.16	4.02	1.29
Group Members Masked	4.40	1.48	6.05	1.07	4.00	1.23
Participant Masked						
Group Members Unmasked	4.64	1.47	6.02	1.11	3.99	1.41
Group Members Masked	4.55	1.45	5.98	1.09	4.02	1.29

Table 2
Results for effects of participant masking and group member masking on IMI competence (H1), IMI choice (H2), and IMI relatedness (H3).

	IMI Competence	IMI Choice	IMI Relatedness
Main Effect Participant Masking			
F	1.801	0.025	0.015
p	0.180	0.874	0.903
η_p^2	0.001	0.000	0.000
Main Effect Group Member Masking			
F	2.867	0.253	0.000
p	0.091	0.615	0.988
η_p^2	0.002	0.000	0.000
Interaction Effect			
F	0.492	1.429	0.087
p	0.483	0.232	0.768
η_p^2	0.000	0.001	0.000

df: Participant Masking=1, Group Member Masking=1, Interaction=1, Error=1341.

6.2. H1–H3: Effect of manipulation on competence, autonomy, relatedness

From Tables 1 and 2, factorial 2 × 2 ANOVAs (participant masking × group member masking) found no main effect of participant masking, no main effect of group member masking, and no interaction effect between participant masking and group member masking on competence (H1.1, H1.2, and H1.3 not supported), autonomy (H2.1, H2.2, and H2.3 not supported), and relatedness (H3.1, H3.2, and H3.3 not supported). Therefore, our manipulation did not have a statistically significant effect on competence, autonomy, or relatedness.

6.3. H4–H8: Analyses of direct and mediated effect

6.3.1. Assumption checks

Assumption checks for mediation analyses can be found in Supplementary Materials.

6.3.2. Hypothesis tests

We tested the mediation model in Fig. 3. We used a 95% bias-corrected confidence interval based on 10,000 bootstrap samples. From Table 3, competence, autonomy, and relatedness do not mediate the effects of voice condition on any outcome variables (H4.4–H4.9, H5.4–H5.9, H6.4–H6.9, H7.4–H7.9, and H8.4–H8.9 not supported). There were direct effects of voice condition on duration of speaking (c') for the comparisons “All unmasked” vs. “Male confederates masked” and “All unmasked” vs. “All masked” (H4.1 supported, H4.2 supported). There were direct effects of voice condition on correctness of responses (c') for the conditions “All unmasked” vs. “Male confederates masked” and “All unmasked” vs. “All masked” (H6.1 supported, H6.2 supported). There were no direct effects (c') of voice condition on speed in responding (H5.1 and H5.2 not supported), stereotype threat scores (H7.1 and H7.2 not supported), and autonomous respect scores (H8.1 and H8.2 not supported). Therefore, we conclude that the voice conditions “Male confederates masked” and “All masked” had a significantly positive effect on duration of speaking compared to “All unmasked”. The voice conditions “Male confederates masked” and “All masked” also had a significantly positive effect on correctness of responses compared to “All unmasked”. Descriptives for each variable are in Table 4.

6.3.3. Interaction effects

Interaction effects for H4–H8 were tested separately using factorial 2 × 2 ANOVAs (participant masking × group member masking).[9] Interaction effects were not significant for speed in responding ($F[1,1341] = 0.064$, $p = 0.801$, $\eta_p^2 = .000$) (H4.3 not supported), duration of speaking ($F[1,1341] = 0.102$, $p = 0.750$, $\eta_p^2 = .000$) (H5.3 not supported), correctness of responses ($F[1,1341] = 0.104$, $p = 0.748$, $\eta_p^2 = .000$) (H6.3 not supported), stereotype threat scores ($F[1,1341] = 0.023$, $p = 0.880$, $\eta_p^2 = .000$) (H7.3 not supported), and autonomous respect scores ($F[1,1341] = 0.223$, $p = 0.637$, $\eta_p^2 = .000$) (H8.3 not supported).

6.3.4. Analysis extension: Speech word count

Stereotype threat has been shown to increase anxiety, which may influence duration of speech *without* actually influencing the number of words spoken—e.g., nervously speaking quickly (Spieler, 2015). Therefore, we performed an additional analysis on actual number of words spoken by each participant during the meeting. This was a small oversight of our preregistration—this is the only analysis not part of the preregistration. The purpose of this analysis is to confirm whether the voice conditions affect not only duration of speaking, but also the actual number of words spoken.

First, a researcher blind to conditions manually listened and recorded the number of words spoken in each participant audio recording. Filler sounds (e.g., “ah”, “uh”, “um”) were excluded (Bortfeld et al., 2001). The researcher made two separate passes over all recordings to ensure an accurate word count. Next, the average number of words spoken per problem was calculated for each participant. This average number of words was used as the outcome variable (Y). We again tested the same mediation model in Fig. 3 with identical parameters to Section 6.3.2. From Table 5, we can see that competence, autonomy, and relatedness did not mediate the effects of voice condition on number of words spoken. However, there was a direct effect of voice condition on number of words spoken (c') for the comparison “All unmasked” vs. “All masked”. Therefore, we conclude that the voice condition “All masked” had a significantly positive effect on number of words spoken. Similar to H4–H8, an interaction effect was tested separately using a factorial 2 × 2 ANOVA (participant masking × group member masking). The interaction effect was not significant ($F[1,1341] = 0.016$, $p = 0.899$, $\eta_p^2 = .000$). Descriptives can be found in Table 4.

7. Discussion

Masking visual identity cues in online avatars (e.g., race for African-American participants in anagram-solving (Lee, 2009), gender for female participants in mathematics (Lee and Nass, 2012)) positively influences problem-solving performance in stereotype-relevant contexts (Lee, 2009; Lee and Nass, 2012). However, it is not known if masking of identity cues in *voice avatars* can have a similarly positive effect.

We conducted a study with four voice-avatar conditions: (1) All unmasked, (2) Male confederates masked, (3) Female participant masked,

[9] See footnote 8.

Table 3

Mediation results with voice condition (X), competence (M_1), autonomy (M_2), relatedness (M_3), and each outcome variable (Y). Regression coefficients a_x ($X \rightarrow M_x$), b_x ($M_x \rightarrow Y$), c' (direct $X \rightarrow Y$), c (total $X \rightarrow Y$), and $a_x b_x$. We use a multicategorical X with 4 conditions: (1) "All unmasked," (2) "Male confederates masked," (3) "Female participant masked," and (4) "All masked." For each outcome variable, the first line presents results for the comparison between (1) and (2), the second line (1) and (3), and the third line (1) and (4). These specific comparisons are done (and not other comparisons) because we use indicator coding, in which all conditions are compared to the reference condition (in this case, "All unmasked" is the reference condition). See Hayes and Preacher (2014) for more details on this approach. All presented effects are unstandardized. Significant results are bold.

Competence			Autonomy		
a_1	b_1	$a_1 b_1$	a_2	b_2	$a_2 b_2$
Duration of Speaking					
−0.190	**2.320*****	−0.440; CI[−1.042, 0.063]	0.103	1.064	0.110; CI[−0.084, 0.408]
0.051	**2.320*****	0.118; CI[−0.390, 0.658]	0.082	1.064	0.087; CI[−0.117, 0.380]
−0.028	**2.320*****	0.256; CI[−0.578, 0.443]	0.040	1.064	0.043; CI[−0.176, 0.296]
Speed in Responding					
−0.190	**9.673*****	−1.833; CI[−4.346, 0.261]	0.103	**7.165*****	0.006; CI[−0.004, 0.019]
0.051	**9.673*****	0.491; CI[−1.707, 2.846]	0.082	**7.165*****	0.005; CI[−0.006, 0.018]
−0.028	**9.673*****	−0.269; CI[−2.429, 1.937]	0.040	**7.165*****	0.002; CI[−0.009, 0.015]
Correctness of Responses					
−0.190	**0.975*****	−0.185; CI[−0.407, 0.036]	0.103	**0.277*****	0.029; CI[-0.017, 0.088]
0.051	**0.975*****	0.050; CI[−0.162, 0.271]	0.082	**0.277*****	0.023; CI[-0.026, 0.080]
−0.028	**0.975*****	−0.027; CI[−0.241, 0.181]	0.040	**0.277*****	0.011; CI[-0.039, 0.064]
Stereotype Threat Scores					
−0.190	−0.023	0.003; CI[−0.006, 0.013]	0.103	−0.054	−0.006; CI[−0.024, 0.006]
0.051	−0.023	−0.001; CI[−0.008, 0.004]	0.082	−0.054	−0.004; CI[−0.022, 0.007]
−0.028	−0.023	0.000; CI[−0.005, 0.007]	0.040	−0.054	−0.002; CI[−0.017, 0.010]
Autonomous Respect Scores					
−0.190	**0.151*****	−0.029; CI[−0.066, 0.003]	0.103	**0.131*****	0.014; CI[−0.008, 0.040]
0.051	**0.151*****	0.008; CI[−0.026, 0.040]	0.082	**0.131*****	0.011; CI[−0.011, 0.036]
−0.028	**0.151*****	−0.004; CI[−0.038, 0.028]	0.040	**0.131*****	0.005; CI[−0.017, 0.031]

Relatedness			Direct Effect	Total Effect
a_3	b_3	$a_3 b_3$	c'	c
Duration of Speaking				
−0.020	0.511	−0.010; CI[−0.166, 0.135]	**4.557***	**4.217***
−0.030	0.511	−0.015; CI[−0.192, 0.131]	1.886	2.076
−0.008	0.511	−0.004; CI[−0.162, 0.156]	**5.628*****	**5.602****
Speed in Responding				
−0.020	−2.076	0.042; CI[−0.638, 0.765]	4.328	3.274
−0.030	−2.076	0.062; CI[−0.652, 0.852]	−0.222	0.918
−0.008	−2.076	0.016; CI[−0.693, 0.729]	8.260	8.295
Correctness of Responses				
−0.020	**−0.723*****	0.015; CI[−0.126, 0.154]	**0.633***	0.491
−0.030	**−0.723*****	0.022; CI[−0.126, 0.174]	0.277	0.371
−0.008	**−0.723*****	0.006; CI[−0.138, 0.149]	**1.007*****	**0.996*****
Stereotype Threat Scores				
−0.020	**−0.095***	0.002; CI[−0.019, 0.023]	−0.092	−0.091
−0.030	**−0.095***	0.003; CI[−0.018, 0.025]	−0.028	−0.031
−0.008	**−0.095***	0.001; CI[−0.020, 0.021]	−0.093	−0.093
Autonomous Respect Scores				
−0.020	**0.447*****	−0.009; CI[−0.094, 0.076]	0.155	0.131
−0.030	**0.447*****	−0.013; CI[−0.105, 0.077]	0.062	0.067
−0.008	**0.447*****	−0.003; CI[−0.091, 0.084]	0.127	0.125

* significant at $p < 0.05$; ** significant at $p < 0.01$; *** significant at $p < 0.005$; significant $a_x b_x$ based on 95% CI.

Table 4

Descriptives for outcomes in H4 through H8 and number of words spoken. Duration of speaking, speed in responding, and number of words spoken are averages per problem.

	Participant Unmasked				Participant Masked			
	Group Members Unmasked		Group Members Masked		Group Members Unmasked		Group Members Masked	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Duration of Speaking (sec.)	36.52	24.57	40.74	25.28	38.59	22.73	42.12	27.44
Speed in Responding (sec.)	82.59	103.52	85.87	149.82	83.51	104.46	90.89	108.89
Correctness of Responses (0–15)	6.48	3.74	6.97	3.69	6.85	3.82	7.48	4.01
Stereotype Threat Scores (1–7)	3.16	1.74	3.07	1.73	3.13	1.74	3.06	1.71
Autonomous Respect Scores (1–7)	3.82	1.47	3.95	1.40	3.89	1.43	3.95	1.40
Number of Words Spoken	30.89	16.93	32.39	16.35	32.68	17.40	33.93	18.17

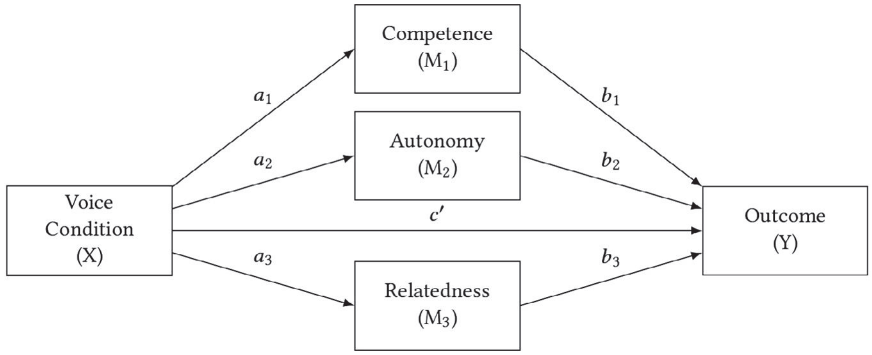

Fig. 3. Mediation model being tested for H4 through H8.

Table 5
Mediation results with voice condition (X), competence (M_1), autonomy (M_2), relatedness (M_3), and number of words spoken (Y). Regression coefficients a_x ($X \rightarrow M_x$), b_x ($M_x \rightarrow Y$), c' (direct $X \rightarrow Y$), c (total $X \rightarrow Y$), and $a_x b_x$. We use a multicategorical X with 4 conditions: (1) "All unmasked," (2) "Male confederates masked," (3) "Female participant masked," and (4) "All masked." For each outcome variable, the first line presents results for the comparison between (1) and (2), the second line (1) and (3), and the third line (1) and (4). These specific comparisons are done (and not other comparisons) because we use indicator coding, in which all conditions are compared to the reference condition (in this case, "All unmasked" is the reference condition). See Hayes and Preacher (2014) for more details on this approach. All presented effects are unstandardized. Significant results are bold.

Competence			Autonomy		
a_1	b_1	$a_1 b_1$	a_2	b_2	$a_2 b_2$
Number of Words Spoken					
−0.190	**3.334*****	−0.632; $CI[-1.397, 0.092]$	0.103	**1.105***	0.114; $CI[-0.073, 0.364]$
0.051	**3.334*****	0.169; $CI[-0.562, 0.910]$	0.082	**1.105***	0.091; $CI[-0.105, 0.326]$
−0.028	**3.334*****	−0.093; $CI[-0.814, 0.635]$	0.040	**1.105***	0.044; $CI[-0.156, 0.266]$

Relatedness			Direct Effect	Total Effect
a_3	b_3	$a_3 b_3$	c'	c
Number of Words Spoken				
−0.020	**−1.73*****	0.002; $CI[-0.018, 0.022]$	1.978	1.494
−0.030	**−1.73*****	0.003; $CI[-0.018, 0.024]$	1.472	1.783
−0.008	**−1.73*****	0.001; $CI[-0.020, 0.021]$	**3.073***	**3.038***

* significant at $p < 0.05$; ** significant at $p < 0.01$; *** significant at $p < 0.005$; significant $a_x b_x$ based on 95% CI.

and (4) All masked. When male confederates used gender-anonymous voices, female participants spoke longer and scored higher on computing problems compared to when no one used gender-anonymous voices. When everyone in the meeting used gender-anonymous voices, female participants spoke longer, spoke more words, and scored higher on computing problems compared to when no one used gender-anonymous voices. However, effects of conditions on subjective measures (competence, choice, relatedness, perceived stereotype threat, and respect from others) and one behavioral measure (speed in responding) were not significant. There were no significant mediation or interaction effects.

Outcomes of our study show that gender-masked voice avatars in online group work have the potential to promote women's participation (i.e., higher duration of speaking and higher spoken word count) and correctness (i.e., higher scores on problems). However, none of these outcomes were significant when female participants alone used a gender-masked voice avatar. Group members using a gender-masked voice avatar resulted in higher duration of speaking and higher problem scores, but not higher spoken word count. Significant positive effects on all three outcomes (speaking duration, spoken word count, problem scores) only occurred when everyone used gender-masked voice avatars. Therefore, while there may be benefits when a subset of meeting participants are using a gender-masked voice avatar, our results suggest that everyone in the meeting should use gender-masked voice avatars to maximize outcomes.[10]

7.1. The benefits of greater participation and performance to diversity

Male-dominated organizations in computing fields can be discouraging for women (e.g., startups with fraternity-like cultures and sexism) (Steinberg, 2014). Repeatedly experiencing stereotype threat can lead to a negative cycle of diminished performance, confidence, and interest (Gilovich et al., 2006). Ultimately, this leads to disidentification from a career (Gilovich et al., 2006). However, our study suggests that voice gender anonymity can have a positive effect on both participation and performance in solving computing problems. Therefore, voice avatars may be one path forward toward improving diversity.

7.2. Applications of identity-masked voice avatars online

The results of this study suggest that identity-masked voice avatars can mitigate stereotype threat in group-based problem solving. When all group members used an identity-masked voice avatar, female participants participated more actively in the online collaborative activity. We view identity-masked voice avatars as applicable for institutions and corporations where meetings between employees are increasingly virtual (Pearlman and Gates, 2010) and where stereotype threat may be present in group decision-making processes wherein stakeholders differ in status (Dubrovsky et al., 1991; Tan et al., 1999). In these instances,

[10] The only statistically significant difference between "Male confederates masked" and "All masked" is spoken word count. However, there is another reason to use voice avatars that mask everyone. In a real virtual group, having only group members' voices gender-masked (from every user's perspective) would create asymmetries in what each user hears. This could be confusing and presents a usability issue.

identity-masked voice avatars (coupled with visual anonymity) may empower more individuals to voice their opinions and share critical information. As a consequence, decision-making processes can be made more robust and foster more equitable discussions.

In addition to applications in online collaborative work and employee meetings, identity-masked voice avatars may also have significant benefits in online learning environments. For instance, while studies within educational games have explored the effects of avatars on educational outcomes (Kao and Harrell, 2018, 2017, 2015), the specific benefits of identity-masked voice avatars remain largely unexplored. Gender bias is rampant in online learning, manifesting in lower peer ratings of female students' work (Morales-Martinez et al., 2020; Brooke, 2021), lower instructor grades of female students' work (Hofer, 2015), and lower student evaluations of female instructors (MacNell et al., 2015; Ayllón, 2022). Gender anonymization within these online learning environments, including identity-masked voice avatars, may help mitigate long-standing gender biases. Within the specific context of computing, the underrepresentation of women may be influenced by a perceived lack of similarity with those in the field (Cheryan and Plaut, 2010). Even subtle gender cues, such as objects in a classroom, can affect female participation (Cheryan et al., 2009). Given that voice carries rich identity cues, identity-masked voice avatars present a promising avenue for exploration in addressing gender inequities in computing.

7.3. Anonymous voice avatars for everyone in the group maximizes benefits

Research has shown that viewing oneself as being in the minority amplifies stereotype threat effects (Lee and Nass, 2012). Within this context, our results suggest that everyone should use gender anonymous voice avatars to maximize outcomes. This minimizes the potential sources of stereotype threat. There is, however, a broader ethical question of whether we *should* use anonymous voice avatars.

7.4. Should we use anonymous voice avatars?

This study may suggest that people should be encouraged to use gender-anonymous avatars in online learning platforms. However, the social implications of anonymizing gender online to combat stereotype threat are fraught—this techno-solutionist approach does not address the underlying cultural problem. Hence, we suggest that this study's results be interpreted as support for a solution that utilizes avatars (both visual and audial) to deemphasize gender cues and instead encourage users to express other aspects of their identities that are more relevant to the context of interaction.

7.5. Anonymous voice avatars did not affect all outcomes

Voice gender anonymity did not have a significant impact on competence, autonomy, relatedness, speed in responding, perceived stereotype threat scores, and autonomous respect scores.

With the exception of speed in responding, these were all self-reported subjective outcomes. Interestingly, even though we see an *actual* difference in levels of competence between conditions (i.e., scores on computing problems), we do *not* see a corresponding difference in self-reported competence. Stereotype threat, through increased rumination, arousal, and efforts to self-regulate, causes cognitive depletion (Beilock, 2008). This cognitive depletion from stereotype threat has been shown to make people poor judges of their own competence (Tellhed and Adolfsson, 2018). As such, one potential explanation is that stereotype threat is undermining participants' assessments. Research into other measures used in our study (i.e., autonomy, relatedness, autonomous respect scores) may shed light on whether the accuracy of other self-reported measures might be inaccurate in some cases. Since SDT was used as a theoretical lens to hypothesize that stereotype threat would have a negative effect on competence, autonomy, and relatedness, future research could utilize alternative measurement approaches for each of these facets to further elucidate the effects of anonymous voice avatars—especially for autonomy and relatedness, for which there were no non-subjective measures in the present study.

Because subjective stereotype threat scores did not differ significantly across conditions, one might argue that stereotype threat was not the cause of differences between conditions. However, given the highly controlled nature of our study and validation of the voice avatar creation software, we believe this is unlikely. Variation between conditions was minimized except for the presence or absence of stereotypical gender cues in the voice-changing output. This was further confirmed through manipulation checks. We believe that a more likely cause is that stereotype threat was a subconscious factor that influenced participants' performance and participation but was not salient enough to manifest in the self-reported scale. The vast majority of previous research on stereotype threat has focused on actual performance effects (Steele, 2010) rather than self-reported measures. Furthermore, previous research has shown that with regards to individuals under stereotype threat conditions, there is a significant divergence between self-reports regarding stereotype threat outcomes and objective stereotype threat outcomes (Bosson et al., 2004). For example, in one study stereotype-threatened individuals demonstrated significantly more anxiety, but this did not manifest as significantly more self-reported anxiety (Bosson et al., 2004). As such, divergences between objective and self-reported stereotype threat measures is supported by previous studies.

Regarding speed in responding, our hypothesis that voice gender anonymity will lead to faster response times may be incorrect. Although we expected participants who are less anxious to also be less tentative (i.e., less hesitant in communicating), it is also possible that such participants would have more confidence and therefore take their time to think carefully about their responses before answering. Study design could have also played a role in this result: Because participant responses were recorded and participants were not engaged in a real-time conversation, this could have encouraged participants to take more time to think about their responses. Here, we did not find any effect of our voice avatar manipulation on speed in responding. Nevertheless, our manipulation had a significant effect on women's participation and performance on computing problems.

7.6. Voice avatar application areas: Education, entertainment, work

There are several potential application areas of voice avatars we envision. In education, this might include teaching through virtual reality (Bailenson et al., 2008), online meeting platforms (de Oliveira Dias et al., 2020), and massive open online courses (Uchidiuno et al., 2016). In entertainment, this might include online multiplayer games. For example, anonymous voice avatars may help reduce toxic behavior between players (Vella et al., 2020; Wadley et al., 2009; Türkay et al., 2020). In work, this might include virtual meetings between employees (Pearlman and Gates, 2010). Although we are optimistic about the potential of voice avatars, there are a few caveats.

Firstly, our results on voice avatars may not generalize to all situations. For example, in a virtual organizational meeting in which the identity of all meeting participants is known in advance, it is not clear what level of effectiveness anonymous voice avatars would have.

Secondly, our meeting platform did not contain a video feed of participants' faces. This absence allowed us to study voice-only avatars in isolation, an important first step. Voice-only avatars are highly applicable in audio calls (e.g., Zoom call-in, regular phone calls, Discord voice chat), social virtual reality, and online gaming. Here, we have shown that the effects of voice avatars alone are impactful. Nevertheless, video feeds are commonplace in many online meetings (Baym et al., 2021).[11] As such, an appropriate next research direction is to study the combination of visual and audial avatars, i.e., visual avatars that align with their audial avatar counterparts, e.g., identity-masked visual avatars.

8. Limitations

Despite the carefully controlled nature of our experiment, there are a few limitations that should be considered for future studies.

Participants were not engaged in live interaction with the other two group members (i.e., the two male confederates). However, according to the manipulation check, participants perceived the interaction as an online meeting. Future smaller-scale studies might consider live interaction.

According to our manipulation checks, participants in self-unmasked conditions still felt their voice was somewhat masked. As a result, the effects of self-masking might be understated in our study (note that our conditions were still successful at inducing higher or lower levels of perceived self-masking—see Manipulation Checks; furthermore, our validation study in Section 3 confirms that the unmasked voice avatar is of a discernible stereotypical gender).

Our study recruited only participants self-identifying as female. More research is needed on non-binary gender identities and anonymous voice avatars. Furthermore, investigating diverse group configurations (smaller vs. larger groups, variations in group demographics) would help researchers understand relevant contextual factors.

9. Conclusion

In a large-scale preregistered study, we investigated gender-anonymous voice avatars in online computing group work. Female participants worked with two ostensible male participants (actually recorded confederates). There were four voice gender anonymity conditions: (1) All unmasked, (2) Male confederates masked, (3) Female participant masked, and (4) All masked. When only male confederates used gender-anonymous voices (compared to all unmasked), female participants spoke for a longer period of time and scored higher on computing problems. When everyone used gender-anonymous voices (compared to all unmasked), female participants spoke for a longer period of time, spoke more words, and scored higher on computing problems. We did not find significant effects on subjective measures and one behavioral measure we deemed to poorly reflect stereotype threat. Our findings demonstrate that anonymous voice avatars are a potential avenue for supporting diversity in virtual group work. We discussed several potential application areas of anonymous voice avatars (e.g., virtual classrooms, online multiplayer games, business meetings). This study presents a first step in understanding the effects of anonymous voice avatars.

CRediT authorship contribution statement

Dominic Kao: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing — review & editing, Supervision, Funding acquisition. **Syed T. Mubarrat:** Conceptualization, Writing – original draft, Writing – review & editing. **Amogh Joshi:** Conceptualization, Writing – original draft, Writing – review & editing. **Swati Pandita:** Conceptualization, Writing – original draft, Writing – review & editing. **Christos Mousas:** Conceptualization, Writing – original draft, Writing – review & editing. **Hai-Ning Liang:** Conceptualization, Writing – original draft, Writing – review & editing. **Rabindra Ratan:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported in part by the National Science Foundation under the award number IIS #2113991. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We would also like to thank the AT&T endowment to the Media & Information Department at MSU, which supports Dr. Ratan's AT&T Chair position.

Appendix A. Supplementary material

Supplementary materials: https://osf.io/ejzcw/.
Preregistration: https://osf.io/zmyhr/.

References

Abbate, Janet, 2012. Recoding Gender: Women's Changing Participation in Computing. MIT Press.

Anon, 2020. Easier access to web pages: ask google assistant to read it aloud. Google, https://blog.google/products/assistant/easier-access-web-pages-let-assistant-read-it-aloud/.

Ansara, Y. Gavriel, Berger, Israel, 2016. Cisgenderism. In: The Wiley Blackwell encyclopedia of gender and sexuality studies. Wiley Online Library, pp. 1–3.

Ayllón, Sara, 2022. Online teaching and gender bias. Econ. Educ. Rev. 89, 102280.

Bailenson, Jeremy N, Yee, Nick, Blascovich, Jim, Beall, Andrew C, Lundblad, Nicole, Jin, Michael, 2008. The use of immersive virtual reality in the learning sciences: digital transformations of teachers, students, and social context. The Journal of the Learning Sciences 17 (1), 102–141.

Balogova, Karolina, Brumby, Duncan, 2022. How do you zoom?: A survey study of how users configure video-conference tools for online meetings. In: 2022 Symposium on Human-Computer Interaction for Work. pp. 1–7.

Bartel, Caroline A., Wrzesniewski, Amy, Wiesenfeld, Batia M., 2012. Knowing where you stand: Physical isolation, perceived respect, and organizational identification among virtual employees. Organ. Sci. 23 (3), 743–757.

Baym, Nancy, Bergmann, Rachel, Coleman, Adam, Fernandez, Ricardo Reyna, Rintel, Sean, Sellen, Abigail, Smith, Tiffany, 2021. Collaboration and meetings - Chapter 1 of the 2021 new future of work report. In: The New Future of Work: Research from Microsoft on the Impact of the Pandemic on Work Practices. Microsoft, https://www.microsoft.com/en-us/research/publication/collaboration-and-meetings/.

Beaton, Ann, Tougas, Francine, Rinfret, Natalie, Huard, Nathalie, Delisle, Marie-Noëlle, 2007. Strength in numbers? Women and mathematics. Eur. J. Psychol. Educ. 22 (3), 291–306.

Beilock, Sian L., 2008. Math performance in stressful situations. Curr. Dir. Psychol. Sci. 17 (5), 339–343.

Beilock, Sian L, Jellison, William A, Rydell, Robert J, McConnell, Allen R, Carr, Thomas H, 2006. On the causal mechanisms of stereotype threat: can skills that don't rely heavily on working memory still be threatened? Pers. Soc. Psychol. Bull. 32 (8), 1059–1071.

Beilock, Sian L., Rydell, Robert J., McConnell, Allen R., 2007. Stereotype threat and working memory: mechanisms, alleviation, and spillover.. J. Exp. Psychol. [Gen.] 136 (2), 256.

Belin, Pascal, Bestelmeyer, Patricia E.G., Latinus, Marianne, Watson, Rebecca, 2011. Understanding voice perception: Understanding voice perception. Br. J. Psychol. 102 (4), 711–725. http://dx.doi.org/10.1111/j.2044-8295.2011.02041.x.

Bell, Amy E, Spencer, Steven J, Iserman, Emma, Logel, Christine ER, 2003. Stereotype threat and women's performance in engineering. J. Eng. Educ. 92 (4), 307–312.

[11] However, people have become increasingly comfortable with turning off their video cameras during online meetings (Baym et al., 2021). This can be due to the stress of having to appear constantly attentive, wanting to multitask, and connectivity issues (Balogova and Brumby, 2022). Furthermore, research has shown that having the video camera on during online meetings increases fatigue and reduces engagement and that such effects are disproportionately worse for women (Shockley et al., 2021).

Bem, Sandra Lipsitz, 1981. Gender schema theory: A cognitive account of sex typing. Psychol. Rev. 88 (4), 354.

Birney, Megan E, Rabinovich, Anna, Morton, Thomas A, Heath, Hannah, Ashcroft, Sam, 2020. When speaking English is not enough: The consequences of language-based stigma for nonnative speakers. J. Lang. Soc. Psychol. 39 (1), 67–86.

Blagojević, Ivan, 2022. Video conferencing statistics. In: 99firms. https://99firms.com/blog/video-conferencing-statistics/.

Blanca, María J., Alarcón, Rafael, Arnau, Jaume, Bono, Roser, Bendayan, Rebecca, 2017. Non-normal data: Is ANOVA still a valid option? Psicothema http://dx.doi.org/10.7334/psicothema2016.383.

Blascovich, Jim, Spencer, Steven J, Quinn, Diane, Steele, Claude, 2001. African Americans and high blood pressure: The role of stereotype threat. Psychol. Sci. 12 (3), 225–229.

Bortfeld, Heather, Leon, Silvia D, Bloom, Jonathan E, Schober, Michael F, Brennan, Susan E, 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. Lang. Speech 44 (2), 123–147.

Bosson, Jennifer K., Haymovitz, Ethan L., Pinel, Elizabeth C., 2004. When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. J. Exp. Soc. Psychol. 40 (2), 247–255.

Boyd, Danah M., Ellison, Nicole B., 2007. Social network sites: Definition, history, and scholarship. J. Comput.-Mediat. Commun. 13 (1), 210–230. http://dx.doi.org/10.1111/j.1083-6101.2007.00393.x.

Brooke, S.J., 2021. Trouble in programmer's paradise: gender-biases in sharing and recognising technical knowledge on stack overflow. Inf. Commun. Soc. 24 (14), 2091–2112.

Brown, Ryan P., Pinel, Elizabeth C., 2003. Stigma on my mind: Individual differences in the experience of stereotype threat. J. Exp. Soc. Psychol. 39 (6), 626–633.

Brynjolfsson, Erik, Horton, John J, Ozimek, Adam, Rock, Daniel, Sharma, Garima, TuYe, Hong-Yi, 2020. COVID-19 and Remote Work: An Early Look at US Data. Technical Report, National Bureau of Economic Research.

Byeon, Hyeon Jeong, Lee, Chaerin, Lee, Jeemin, Oh, Uran, 2022. "A voice that suits the situation": Understanding the needs and challenges for supporting end-user voice customization. In: CHI Conference on Human Factors in Computing Systems. ACM, New Orleans LA USA, pp. 1–10. http://dx.doi.org/10.1145/3491102.3501856.

Chalabaev, Aïna, Sarrazin, Philippe, Stone, Jeff, Cury, François, 2008. Do achievement goals mediate stereotype threat?: An investigation on females' soccer performance. J. Sport Exerc. Psychol. 30 (2), 143–158.

Chan, Brian, Gray, Kishonna, 2020. Microstreaming, microcelebrity, and marginalized masculinity: Pathways to visibility and self-definition for black men in gaming. Women's Stud. Commun. 43 (4), 354–362. http://dx.doi.org/10.1080/07491409.2020.1833634.

Chang, Felix, Luo, Mufan, Walton, Gregory, Aguilar, Lauren, Bailenson, Jeremy, 2019. Stereotype threat in virtual learning environments: Effects of avatar gender and sexist behavior on women's math learning outcomes. Cyberpsychology Behav. Soc. Netw. 22 (10), 634–640.

Cherif, Emna, Lemoine, Jean-François, 2017. Human vs. synthetic recommendation agents' voice: The effects on consumer reactions. In: Rossi, Patricia (Ed.), Marketing At the Confluence Between Entertainment and Analytics. In: Developments in Marketing Science: Proceedings of the Academy of Marketing Science, Springer International Publishing, Cham, pp. 301–310. http://dx.doi.org/10.1007/978-3-319-47331-4_53.

Cheryan, Sapna, Meltzoff, Andrew N., Kim, Saenam, 2011. Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. Comput. Educ. 57 (2), 1825–1835.

Cheryan, Sapna, Plaut, Victoria C., 2010. Explaining underrepresentation: A theory of precluded interest. Sex Roles 63 (7), 475–488.

Cheryan, S., Plaut, V.C., Davies, P., Steele, C.M., 2009. Ambient belonging: How stereotypical environments impact gender participation in computer science. J. Personal. Soc. Psychol. 97.

Clark, Leigh, Doyle, Philip, Garaialde, Diego, Gilmartin, Emer, Schlögl, Stephan, Edlund, Jens, Aylett, Matthew, Cabral, João, Munteanu, Cosmin, Edwards, Justin, R Cowan, Benjamin, 2019. The state of speech in HCI: Trends, themes and challenges. Interact. Comput. 31 (4), 349–371. http://dx.doi.org/10.1093/iwc/iwz016.

Cohen, Geoffrey L., Garcia, Julio, Apfel, Nancy, Master, Allison, 2006. Reducing the racial achievement gap: A social-psychological intervention. Science 313 (5791), 1307–1310. http://dx.doi.org/10.1126/science.1128317.

Cohen, Geoffrey L., Garcia, Julio, Purdie-Vaughns, Valerie, Apfel, Nancy, Brzustoski, Patricia, 2009. Recursive processes in self-affirmation: Intervening to close the minority achievement gap. Science 324 (5925), 400–403. http://dx.doi.org/10.1126/science.1170769.

Cohen, Geoffrey L., Purdie-Vaughns, Valerie, Garcia, Julio, 2012. An Identity Threat Perspective on Intervention. Oxford University Press.

Cohoon, J. McGrath, Aspray, William, 2006. Women and Information Technology: Research on Underrepresentation. The MIT Press.

Cote, Amanda C., 2017. "I can defend myself" women's strategies for coping with harassment while gaming online. Games Cult. 12 (2), 136–155.

Croizet, Jean-Claude, Claire, Theresa, 1998. Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. Pers. Soc. Psychol. Bull. 24 (6), 588–594.

Danaher, Kelly, Crandall, Christian S., 2008. Stereotype threat in applied settings re-examined. J. Appl. Soc. Psychol. 38 (6), 1639–1655.

Davies, Paul G., Spencer, Steven J., Steele, Claude M., 2005. Clearing the air: identity safety moderates the effects of stereotype threat on women's leadership aspirations. J. Pers. Soc. Psychol. 88 (2), 276.

de Oliveira Dias, Murillo, Lopes, RD.O.A., Teles, Andre Correia, 2020. Will virtual replace classroom teaching? Lessons from virtual classes via zoom in the times of COVID-19. J. Adv. Educ. Philos. 4 (05), 208–213.

Dean, Cornelia, 2007. Computer science takes steps to bring women to the fold. N. Y. Times 17, https://www.nytimes.com/2007/04/17/science/17comp.html.

Deci, Edward L., Ryan, Richard M., 1987. The support of autonomy and the control of behavior. J. Pers. Soc. Psychol. 53 (6), 1024.

Delgado, Ana R., Prieto, Gerardo, 2008. Stereotype threat as validity threat: The anxiety–sex–threat interaction. Intelligence 36 (6), 635–640.

Divekar, Rahul R, Kephart, Jeffrey O, Mou, Xiangyang, Chen, Lisha, Su, Hui, 2019a. You talkin'to me? A practical attention-aware embodied agent. In: IFIP Conference on Human-Computer Interaction. Springer, pp. 760–780.

Divekar, Rahul R, Mou, Xiangyang, Chen, Lisha, De Bayser, Maira Gatti, Guerra, Melina Alberio, Su, Hui, 2019b. Embodied conversational AI agents in a multi-modal multi-agent competitive dialogue. In: IJCAI. pp. 6512–6514.

Dubrovsky, Vitaly J., Kiesler, Sara, Sethna, Beheruz N., 1991. The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. Hum.-Comput. Interact. 6 (2), 119–146. http://dx.doi.org/10.1207/s15327051hci0602_2.

Ellison, Glenn, Swanson, Ashley, 2010. The gender gap in secondary school mathematics at high achievement levels: Evidence from the American mathematics competitions. J. Econ. Perspect. 24 (2), 109–128.

Eysenck, Michael W., Calvo, Manuel G., 1992. Anxiety and performance: The processing efficiency theory. Cogn. Emot. 6 (6), 409–434.

Fauville, Geraldine, Luo, Mufan, Muller Queiroz, Anna Carolina, Bailenson, Jeremy N, Hancock, Jeff, 2021. Nonverbal mechanisms predict zoom fatigue and explain why women experience higher levels than men. Available at SSRN 3820035.

Ferreira, Rafael, Pereira, Ruben, Bianchi, Isaías Scalabrin, da Silva, Miguel Mira, 2021. Decision factors for remote work adoption: Advantages, disadvantages, driving forces and challenges. J. Open Innov. Technol. Mark. Complexity 7 (1), 1–24. http://dx.doi.org/10.3390/joitmc7010070.

Fordham, J., Ratan, R., Huang, K.T., Zhou, W., Silva, K., 2020. Is Gender Disparity in STEM Fields Related to Gender Stereotypes in Videogames? An Experimental Examination of Stereotype Threat Context Transfer. The American Behavioral Scientist, Newbury Park, CA.

Gambino, Andrew, Fox, Jesse, Ratan, Rabindra A., 2020. Building a stronger CASA: Extending the computers are social actors paradigm. Hum.-Mach. Commun. 1, 71–85.

Garcia, Julio, Cohen, Geoffrey L., 2013. A social psychological approach to educational intervention. Behav. Found. Public Policy 329.

Gaucher, Danielle, Friesen, Justin, Kay, Aaron C., 2011. Evidence that gendered wording in job advertisements exists and sustains gender inequality. J. Pers. Soc. Psychol. 101 (1), 109.

Giammarco, Erica A, Schneider, Travis J, Carswell, Julie J, Knipe, William S, 2015. Video game preferences and their relation to career interests. Pers. Individ. Differ. 73, 98–104.

Gilovich, Thomas, Keltner, Dacher, Nisbett, Richard E, 2006. Being a member of a stigmatized group: stereotype threat. Soc. Psychol. 467–468.

Goberman, Alexander M., Hughes, Stephanie, Haydock, Todd, 2011. Acoustic characteristics of public speaking: Anxiety and practice effects. Speech Commun. 53 (6), 867–876.

Goldsmith, Pat Antonio, 2004. Schools' racial mix, students' optimism, and the black-white and latino-white achievement gaps. Sociol. Educ. 77 (2), 121–147.

Gonzales, Patricia M., Blanton, Hart, Williams, Kevin J., 2002. The effects of stereotype threat and double-minority status on the test performance of latino women. Pers. Soc. Psychol. Bull. 28 (5), 659–670.

Gratton, Lynda, Kelan, Elisabeth, Voigt, Andreas, Walker, Lamia, Wolfram, Hans-Joachim, 2007. Innovative Potential: Men and Women in Teams. The Lehman Brothers Centre for Women in Business. London Business School.

Gray, Kishonna L., 2012. Intersecting oppressions and online communities: Examining the experiences of women of color in Xbox Live. Inf. Commun. Soc. 15 (3), 411–428. http://dx.doi.org/10.1080/1369118X.2011.642401.

Gupta, Vishal K., Bhawe, Nachiket M., 2007. The influence of proactive personality and stereotype threat on women's entrepreneurial intentions. J. Leadersh. Organ. Stud. 13 (4), 73–85.

Hargittai, Eszter, Shafer, Steven, 2006. Differences in actual and perceived online skills: The role of gender. Soc. Sci. Q. 87 (2), 432–448.

Harter, Susan, 1981. A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. Dev. Psychol. 17 (3), 300–312. http://dx.doi.org/10.1037/0012-1649.17.3.300.

Hasher, Lynn, Zacks, Rose T., 1988. Working memory, comprehension, and aging: A review and a new view. Psychol. Learn. Motiv. 22, 193–225.

Hayes, Andrew F., 2022. Introduction to Mediation, Moderation, and Conditional Process Analysis, third ed.

Hayes, Andrew F., Preacher, Kristopher J., 2014. Statistical mediation analysis with a multicategorical independent variable. Br. J. Math. Stat. Psychol. 67 (3), 451–470. http://dx.doi.org/10.1111/bmsp.12028.

Henn, Steve, 2014. When women stopped coding. In: NPR. https://www.npr.org/sections/money/2014/10/21/357629765/when-women-stopped-coding.

Hess, Thomas M., Hinson, Joey T., 2006. Age-related variation in the influences of aging stereotypes on memory in adulthood. Psychol. Aging 21 (3), 621.

Higher Education Statistics Agency, 2018. Higher education student statistics: UK, 2016/17 - Summary. In: HESA. https://www.hesa.ac.uk/news/11-01-2018/sfr247-higher-education-studentstatistics.

Hofer, Sarah I., 2015. Studying gender bias in physics grading: The role of teaching experience and country. Int. J. Sci. Educ. 37 (17), 2879–2905.

Hoyt, Crystal L., Murphy, Susan E., 2016. Managing to clear the air: Stereotype threat, women, and leadership. Leadersh. Q. 27 (3), 387–399.

Inzlicht, Michael, Ben-Zeev, Talia, 2000. A threatening intellectual environment: When and why females are susceptible to experiencing problem-solving deficits in the presence of males. Psychol. Sci. 1, http://www.il.proquest.com/umi/; http://search.epnet.com/login.aspx?direct=true{&}db=psyh{&}an=2001-95016-421.

Inzlicht, Michael, Ben-Zeev, Talia, 2003. Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. J. Educ. Psychol. 95 (4), 796.

Inzlicht, Michael, Good, Catherine, 2006. How environments can threaten academic performance, self-knowledge, and sense of belonging. In: Stigma and Group Inequality. Psychology Press, pp. 143–164.

Johns, Michael, Inzlicht, Michael, Schmader, Toni, 2008. Stereotype threat and executive resource depletion: examining the influence of emotion regulation. J. Exp. Psychol. [Gen.] 137 (4), 691.

Johns, Michael, Schmader, Toni, Martens, Andy, 2005. Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. Psychol. Sci. 16 (3), 175–179.

Joint Council for Qualifications, 2018. GCSEs 2018. JCQ, https://www.jcq.org.uk/wpcontent/uploads/2018/12/.

Kao, Dominic, 2019. The effects of anthropomorphic avatars vs. non-anthropomorphic avatars in a jumping game. In: Proceedings of the 14th international conference on the foundations of digital games. pp. 1–5.

Kao, Dominic, Harrell, D Fox, 2015. Exploring the impact of role model avatars on game experience in educational games. In: Proceedings of the 2015 annual symposium on computer-human interaction in play. pp. 571–576.

Kao, Dominic, Harrell, D Fox, 2017. Mazestar: a platform for studying virtual identity and computer science education. In: Proceedings of the 12th international conference on the foundations of digital games. pp. 1–6.

Kao, Dominic, Harrell, D Fox, 2018. The effects of badges and avatar identification on play and making in educational games. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–19.

Kao, Dominic, Ratan, Rabindra, Mousas, Christos, Joshi, Amogh, Melcer, Edward F, 2022. Audio matters too: How audial avatar customization enhances visual avatar customization. CHI Conference on Human Factors in Computing Systems. CHI '22, April 29–May 5, 2022, New Orleans, la, USA, Association for Computing Machinery, http://dx.doi.org/10.1145/3491102.3501848, arXiv:4159467.

Kao, Dominic, Ratan, Rabindra, Mousas, Christos, Magana, Alejandra, 2021. The effects of a self-similar avatar voice in educational games. In: CHI PLAY.

Karl, Katherine A., Peluchette, Joy V., Aghakhani, Navid, 2022. Virtual work meetings during the COVID-19 pandemic: The good, bad, and ugly. Small Group Res. 53 (3), 343–365.

Kaye, Linda K., Pennington, Charlotte R., McCann, Joseph J., 2018. Do casual gaming environments evoke stereotype threat? Examining the effects of explicit priming and avatar gender. Comput. Hum. Behav. 78, 142–150.

Keller, Johannes, 2002. Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. Sex Roles 47 (3), 193–198.

Khan, Anam Ahmad, Nawaz, Sadia, Newn, Joshua, Kelly, Ryan M, Lodge, Jason M, Bailey, James, Velloso, Eduardo, 2022. To type or to speak? The effect of input modality on text understanding during note-taking. In: CHI Conference on Human Factors in Computing Systems. pp. 1–15.

Ko, Sei Jin, Judd, Charles M., Blair, Irene V., 2006. What the voice reveals: Within-and between-category stereotyping on the basis of voice. Pers. Soc. Psychol. Bull. 32 (6), 806–819. http://dx.doi.org/10.1177/0146167206286627.

Koch, Sabine C., Müller, Stephanie M., Sieverding, Monika, 2008. Women and computers. Effects of stereotype threat on attribution of failure. Comput. Educ. 51 (4), 1795–1803.

Kray, Laura J., Galinsky, Adam D., Thompson, Leigh, 2002. Reversing the gender gap in negotiations: An exploration of stereotype regeneration. Organ. Behav. Hum. Decis. Process. 87 (2), 386–409.

Kuznekoff, Jeffrey H., Rose, Lindsey M., 2013. Communication in multiplayer gaming: Examining player responses to gender cues. New Media Soc. 15 (4), 541–556.

Laukka, Petri, Linnman, Clas, Åhs, Fredrik, Pissiota, Anna, Frans, Örjan, Faria, Vanda, Michelgård, Åsa, Appel, Lieuwe, Fredrikson, Mats, Furmark, Tomas, 2008. In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. J. Nonverbal Behav. 32 (4), 195–214.

Lee, Jong-Eun Roselyn, 2009. To reveal or to cloak? Effects of identity salience on stereotype threat responses in avatar-represented group contexts. Int. J. Internet Sci. 4 (1), 34–49, http://proxy-remote.galib.uga.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true{&}db=a9h{&}AN=54306690{&}site=ehost-live.

Lee, Jong-Eun Roselyn, Nass, Clifford, 2012. Distinctiveness-based stereotype threat and the moderating role of coaction contexts. J. Exp. Soc. Psychol. 48 (1), 192–199. http://dx.doi.org/10.1016/j.jesp.2011.06.018.

Lee, Jong-Eun Roselyn, Nass, Clifford I., Bailenson, Jeremy N., 2014. Does the mask govern the mind?: Effects of arbitrary gender representation on quantitative task performance in avatar-represented virtual groups. Cyberpsychology Behav. Soc. Netw. 17 (4), 248–254.

Levy, Becca, 1996. Improving memory in old age through implicit self-stereotyping. J. Pers. Soc. Psychol. 71 (6), 1092.

Lips, Hilary M., 2017. Sandra bem: Naming the impact of gendered categories and identities. Sex Roles 76 (9), 627–632.

Logel, Christine, Walton, Gregory M, Spencer, Steven J, Iserman, Emma C, von Hippel, William, Bell, Amy E, 2009. Interacting with sexist men triggers social identity threat among female engineers. J. Pers. Soc. Psychol. 96 (6), 1089.

Maass, Anne, D'Ettole, Claudio, Cadinu, Mara, 2008. Checkmate? The role of gender stereotypes in the ultimate intellectual sport. Eur. J. Soc. Psychol. 38 (2), 231–245.

MacNell, Lillian, Driscoll, Adam, Hunt, Andrea N., 2015. What's in a name: Exposing gender bias in student ratings of teaching. Innov. High. Educ. 40, 291–303.

Martens, Andy, Johns, Michael, Greenberg, Jeff, Schimel, Jeff, 2006. Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. J. Exp. Soc. Psychol. 42 (2), 236–243.

Master, Allison, Cheryan, Sapna, Meltzoff, Andrew N., 2016. Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. J. Educ. Psychol. 108 (3), 424–437.

McAleer, Phil, Todorov, Alexander, Belin, Pascal, 2014. How do you say 'hello'? Personality impressions from brief novel voices. PLoS One 9 (3), e90779. http://dx.doi.org/10.1371/journal.pone.0090779.

McAuley, Edward, Duncan, Terry, Tammen, Vance V., 1989. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. Res. Q. Exerc. Sport 60 (1), 48–58.

McGlone, Matthew S., Aronson, Joshua, 2006. Stereotype threat, identity salience, and spatial reasoning. J. Appl. Dev. Psychol. 27 (5), 486–493.

McGlone, Matthew S., Aronson, Joshua, Kobrynowicz, Diane, 2006. Stereotype threat and the gender gap in political knowledge. Psychol. Women Q. 30 (4), 392–398.

McGlone, Matthew S., Pfiester, R. Abigail, 2015. Stereotype threat and the evaluative context of communication. J. Lang. Soc. Psychol. 34 (2), 111–137.

McIntyre, Rusty B., Paulson, René M., Lord, Charles G., 2003. Alleviating women's mathematics stereotype threat through salience of group achievements. J. Exp. Soc. Psychol. 39 (1), 83–90.

McLean, Lavinia, Griffiths, Mark D., 2013. Female gamers: A thematic analysis of their gaming experience. Int. J. Game-Based Learn. (IJGBL) 3 (3), 54–71.

McLean, Lavinia, Griffiths, Mark D., 2019. Female gamers' experience of online harassment and social support in online gaming: A qualitative study. Int. J. Ment. Health Addict. 17 (4), 970–994.

Mendoza-Denton, Rodolfo, Page-Gould, Elizabeth, 2008. Can cross-group friendships influence minority students' well-being at historically white universities? Psychol. Sci. 19 (9), 933–939.

Miyazoe, Terumi, Anderson, Terry, 2011. Voice interaction online. p. 29.

Morales-Martinez, Gabriela, Latreille, Paul, Denny, Paul, 2020. Nationality and gender biases in multicultural online learning environments: The effects of anonymity. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–14.

Murphy, Mary C., Steele, Claude M., Gross, James J., 2007. Signaling threat: How situational cues affect women in math, science, and engineering settings. Psychol. Sci. 18 (10), 879–885.

Murphy, Mary C., Taylor, Valerie Jones, 2012. The Role of Situational Cues in Signaling and Maintaining Stereotype Threat. Oxford University Press.

Nass, Clifford, Moon, Youngme, 2000. Machines and mindlessness: Social responses to computers. J. Soc. Issues 56 (1), 81–103.

Nguyen, Hannah-Hanh D., Ryan, Ann Marie, 2008. Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. J. Appl. Psychol. 93 (6), 1314.

Nowak, Kristine L., Fox, Jesse, 2018. Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. Rev. Commun. Res. 6, 30–53.

O'Brien, Laurie T., Crandall, Christian S., 2003. Stereotype threat and arousal: Effects on women's math performance. Pers. Soc. Psychol. Bull. 29 (6), 782–789.

Okano, Hiiro, Mizuno, Keisuke, Miyakawa, Haruna, Zempo, Keiichi, 2022. Avatar voice morphing to match subjective and objective self voice perception. In: Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology. pp. 1–2.

Osborne, Jason W., 2001. Testing stereotype threat: Does anxiety explain race and sex differences in achievement? Contemp. Educ. Psychol. 26 (3), 291–310.

Osborne, Jason W., 2007. Linking stereotype threat and anxiety. Educ. Psychol. 27 (1), 135–154.

Palomares, Nicholas A., 2008. Explaining gender-based language use: Effects of gender identity salience on references to emotion and tentative language in intra-and intergroup contexts. Hum. Commun. Res. 34 (2), 263–286.

Palomares, Nicholas A., 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. Commun. Res. 36 (4), 538–560.

Parker, Kim, Lenhart, Amanda, 2011. I. Online Learning. Pew Research Center's Social & Demographic Trends Project.

Paule-Ruiz, MPuerto, Álvarez-García, Víctor, Pérez-Pérez, J.R., Riestra-González, M., 2013. Voice interactive learning: A framework and evaluation. In: Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education. In: ITiCSE '13, Association for Computing Machinery, New York, NY, USA, pp. 34–39. http://dx.doi.org/10.1145/2462476.2462489.

Pearlman, David M., Gates, Nicholas A., 2010. Hosting business meetings and special events in virtual worlds: a fad or the future? J. Conv. Event Tour. 11 (4), 247–265.

Peck, Tabitha C., Good, Jessica J., Bourne, Kimberly A., 2020. Inducing and mitigating stereotype threat through gendered virtual body-swap illusions. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–13.

Pegg, Karlee J., O'Donnell, Alexander W., Lala, Girish, Barber, Bonnie L., 2018. The role of online social identity in the relationship between alcohol-related content on social networking sites and adolescent alcohol use. Cyberpsychology Behav. Soc. Netw. 21 (1), 50–55. http://dx.doi.org/10.1089/cyber.2016.0665.

Pennington, Charlotte R, Heim, Derek, Levy, Andrew R, Larkin, Derek T, 2016. Twenty years of stereotype threat research: A review of psychological mediators. PLoS One 11 (1), e0146487.

Ratan, Rabindra, Beyea, David, Li, Benjamin J., Graciano, Luis, 2020. Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the Proteus effect. Media Psychol. 23 (5), 651–675. http://dx.doi.org/10.1080/15213269.2019.1623698.

Ratan, Rabindra, Miller, Dave B., Bailenson, Jeremy N., 2022. Facial appearance dissatisfaction explains differences in zoom fatigue. Cyberpsychology Behav. Soc. Netw. 25 (2), 124–129.

Ratan, Rabindra, Sah, Young June, 2015. Leveling up on stereotype threat: The role of avatar customization and avatar embodiment. Comput. Hum. Behav. 50, 367–374. http://dx.doi.org/10.1016/j.chb.2015.04.010.

Reeves, Byron, Nass, Clifford, 1996. The media equation: How people treat computers, television, and new media like real people, 10. 236605, Cambridge, UK.

Renub Research, 2022. Online Education Market, Size, Global Forecast 2022–2027, Industry Trends, Share, Growth, Impact of COVID-19, Opportunity Company Analysis. Technical Report.

Rosenthal, Harriet E.S., Crisp, Richard J., 2006. Reducing stereotype threat by blurring intergroup boundaries. Pers. Soc. Psychol. Bull. 32 (4), 501–511.

Rosenthal, Harriet E.S., Crisp, Richard J., Suen, Mein-Woei, 2007. Improving performance expectancies in stereotypic domains: Task relevance and the reduction of stereotype threat. Eur. J. Soc. Psychol. 37 (3), 586–597.

Ryan, R.M., Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. 55 (1), 68–78. http://dx.doi.org/10.1037/0003-066X.55.1.68, arXiv:0208024 [gr-qc].

Ryan, Richard M., Deci, Edward L., 2020. Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. Contemp. Educ. Psychol. 61, 101860. http://dx.doi.org/10.1016/j.cedpsych.2020.101860.

Sax, Linda J., Lehman, Kathleen J., Jacobs, Jerry A., Kanny, M. Allison, Lim, Gloria, Monje-Paulson, Laura, Zimmerman, Hilary B., 2017. Anatomy of an enduring gender gap: The evolution of women's participation in computer science. J. Higher Educ. 88 (2), 258–293. http://dx.doi.org/10.1080/00221546.2016.1257306.

Schmader, Toni, 2002. Gender identification moderates stereotype threat effects on women's math performance. J. Exp. Soc. Psychol. 38 (2), 194–201.

Schmader, Toni, 2012. Stereotype Threat: Theory, Process, and Application. Oxford University Press.

Schmader, Toni, Forbes, Chad E, Zhang, Shen, Berry Mendes, Wendy, 2009. A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. Pers. Soc. Psychol. Bull. 35 (5), 584–596.

Schmader, Toni, Johns, Michael, 2003. Converging evidence that stereotype threat reduces working memory capacity. J. Pers. Soc. Psychol. 85 (3), 440.

Schmader, Toni, Johns, Michael, Forbes, Chad, 2008. An integrated process model of stereotype threat effects on performance. Psychol. Rev. 115 (2), 336–356. http://dx.doi.org/10.1037/0033-295X.115.2.336.

Schoemann, Alexander M., Boulton, Aaron J., Short, Stephen D., 2017. Determining power and sample size for simple and complex mediation models. Soc. Psychol. Pers. Sci. 8 (4), 379–386. http://dx.doi.org/10.1177/1948550617715068.

Seaborn, Katie, Miyake, Norihisa P., Pennefather, Peter, Otake-Matsuura, Mihoko, 2022. Voice in human–agent interaction: A survey. ACM Comput. Surv. 54 (4), 1–43. http://dx.doi.org/10.1145/3386867.

Sebastian, Richard J., Bouchard Ryan, Ellen, 2018. Speech cues and social evaluation: Markers of ethnicity, social class, and age. In: Giles, Howard, St. Clair, Robert N. (Eds.), Recent Advances in Language, Communication, and Social Psychology, first ed. Routledge, pp. 112–143. http://dx.doi.org/10.4324/9780429436178-5.

Sekaquaptewa, Denise, Thompson, Mischa, 2003. Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. J. Exp. Soc. Psychol. 39 (1), 68–74.

Shah, Harsh D., Sundas, Amit, Sharma, Shabnam, 2021. Controlling email system using audio with speech recognition and text to speech. In: 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions). ICRITO, pp. 1–7. http://dx.doi.org/10.1109/ICRITO51393.2021.9596293.

Shapiro, Jenessa R., Neuberg, Steven L., 2007. From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. Pers. Soc. Psychol. Rev. 11 (2), 107–130.

Shih, Margaret, Pittinsky, Todd L., Ambady, Nalini, 1999. Stereotype susceptibility: Identity salience and shifts in quantitative performance. Psychol. Sci. 10 (1), 80–83.

Shockley, Kristen M, Gabriel, Allison S, Robertson, Daron, Rosen, Christopher C, Chawla, Nitya, Ganster, Mahira L, Ezerins, Maira E, 2021. The fatiguing effects of camera use in virtual meetings: A within-person field experiment. J. Appl. Psychol. 106 (8), 1137.

Smith, Morgan, 2022. Twitter, reddit and 8 other companies offering permanent remote or hybrid work—and hiring right now. In: CNBC. https://www.cnbc.com/2022/04/13/10-companies-that-switched-to-permanent-hybrid-or-remote-work-and-hiring-right-now.html.

Smith, Jessi L, Lewis, Karyn L, Hawthorne, Lauren, Hodges, Sara D, 2013. When trying hard isn't natural: Women's belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. Pers. Soc. Psychol. Bull. 39 (2), 131–143.

Smith, Jessi L., Sansone, Carol, White, Paul H., 2007. The stereotyped task engagement process: The role of interest and achievement motivation. J. Educ. Psychol. 99 (1), 99.

Spencer, Steven J., Steele, Claude M., Quinn, Diane M., 1999. Stereotype threat and women's math performance. J. Exp. Soc. Psychol. 35 (1), 4–28.

Spieler, Claire A., 2015. Using Awareness Training to Decrease Nervous Habits in Public Speaking. University of South Florida.

Sproull, Lee, Kiesler, Sara, Kiesler, Sara B., 1991. Connections: New Ways of Working in the Networked Organization. MIT Press.

Steele, Claude M., 1997. A threat in the air: How stereotypes shape intellectual identity and performance. Am. Psychol. 52 (6), 613.

Steele, Claude M., 2010. Whistling vivaldi and other clues to how stereotypes affect us. In: Whistling Vivaldi. http://dx.doi.org/10.1126/science.1194619.

Steele, Claude M., Aronson, Joshua, 1995. Stereotype threat and the intellectual test performance of African Americans. J. Personal. Soc. Psychol. 69 (5), 797–811. http://dx.doi.org/10.1037/0022-3514.69.5.797.

Steele, Jennifer, James, Jacquelyn B., Barnett, Rosalind Chait, 2002a. Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. Psychol. Women Q. 26 (1), 46–50.

Steele, Claude M., Spencer, Steven J., Aronson, Joshua, 2002b. Contending with group image: The psychology of stereotype and social identity threat. In: Advances in Experimental Social Psychology, vol. 34. Elsevier, pp. 379–440. http://dx.doi.org/10.1016/S0065-2601(02)80009-0.

Steinberg, Joseph, 2014. Sexism in startups: The frank conversation we need to be having. In: Forbes. https://www.forbes.com/sites/josephsteinberg/2014/09/18/heres-the-conversation-men-need-to-have-about-sexism-as-told-by-a-man/?sh=bf61c41fc398.

Stone, Jeff, Chalabaev, Aina, Harrison, C. Keith, 2012. The Impact of Stereotype Threat on Performance in Sports. Oxford University Press.

Stone, Jeff, McWhinnie, Chad, 2008. Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. J. Exp. Soc. Psychol. 44 (2), 445–452.

Stout, Jane G., Dasgupta, Nilanjana, 2011. When he doesn't mean you: Gender-exclusive language as ostracism. Pers. Soc. Psychol. Bull. 37 (6), 757–769.

Strand, Elizabeth A., 1999. Uncovering the role of gender stereotypes in speech perception. J. Lang. Soc. Psychol. 18 (1), 86–100. http://dx.doi.org/10.1177/0261927X99018001006.

Tan, Bernard C.Y., Wei, Kwok-kee, Watson, Richard T., 1999. The equalizing impact of a group support system on status differentials. ACM Trans. Inf. Syst. 17 (1), 77–100. http://dx.doi.org/10.1145/297117.297126.

Taylor, Valerie Jones, Walton, Gregory M., 2011. Stereotype threat undermines academic learning. Pers. Soc. Psychol. Bull. 37 (8), 1055–1067.

Tellhed, Una, Adolfsson, Caroline, 2018. Competence and confusion: How stereotype threat can make you a bad judge of your competence. Eur. J. Soc. Psychol. 48 (2), O189–O197.

Thoman, Dustin B., Smith, Jessi L., Brown, Elizabeth R., Chase, Justin, Lee, Joo Young K., 2013. Beyond performance: A motivational experiences model of stereotype threat. Educ. Psychol. Rev.

Trent, Sonja A., 1995. Voice quality: Listener identification of African-American versus Caucasian speakers. J. Acoust. Soc. Am. 98 (5), 2936. http://dx.doi.org/10.1121/1.414099.

Türkay, Selen, Formosa, Jessica, Adinolf, Sonam, Cuthbert, Robert, Altizer, Roger, 2020. See no evil, hear no evil, speak no evil: How collegiate players define, experience and cope with toxicity. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–13.

Turner, Laure, 2009. Gender diversity and innovative performance. Int. J. Innov. Sustain. Dev. 4 (2–3), 123–134.

Tyler, Tom R., Blader, Steven L., 2002. Autonomous vs. comparative status: Must we be better than others to feel good about ourselves? Organ. Behav. Hum. Decis. Process. 89 (1), 813–838. http://dx.doi.org/10.1016/S0749-5978(02)00031-6.

Uchidiuno, Judith, Ogan, Amy, Yarzebinski, Evelyn, Hammer, Jessica, 2016. Understanding ESL students' motivations to increase MOOC accessibility. In: Proceedings of the Third (2016) ACM Conference on Learning@ Scale. pp. 169–172.

Vella, Kellie, Klarkowski, Madison, Turkay, Selen, Johnson, Daniel, 2020. Making friends in online games: gender differences and designing for greater social connectedness. Behav. Inf. Technol. http://dx.doi.org/10.1080/0144929X.2019.1625442.

Wadley, Greg, Gibbs, Martin R., Ducheneaut, Nicolas, 2009. You can be too rich: Mediated communication in a virtual world. In: Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group - Design: Open 24/7. OZCHI '09, http://dx.doi.org/10.1145/1738826.1738835.

Wallace, Kathleen A., 1999. Anonymity. Ethics Inf. Technol. 1 (1), 21–31.

Wallace, Kathleen A., 2008. Online anonymity. In: The Handbook of Information and Computer Ethics. Wiley Online Library, p. 165.

Walther, Joseph B., 1996. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. Commun. Res. 23 (1), 3–43.

Walton, Gregory M., Cohen, Geoffrey L., 2007. A question of belonging: Race, social fit, and achievement. J. Personal. Soc. Psychol. 92 (1), 82–96. http://dx.doi.org/10.1037/0022-3514.92.1.82.

Walton, Gregory M., Spencer, Steven J., 2009. Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. Psychol. Sci. 20 (9), 1132–1139.

Woolley, Anita Williams, Chabris, Christopher F, Pentland, Alex, Hashmi, Nada, Malone, Thomas W, 2010. Evidence for a collective intelligence factor in the performance of human groups. Science 330 (6004), 686–688.

Wörtwein, Torsten, Morency, Louis-Philippe, Scherer, Stefan, 2015. Automatic assessment and analysis of public speaking anxiety: A virtual audience case study. In: 2015 International Conference on Affective Computing and Intelligent Interaction. ACII, IEEE, pp. 187–193.

Wu, Juncheng, Rajesh, Anushka, Huang, Yu-Ning, Chhugani, Karishma, Acharya, Rajesh, Peng, Kerui, Johnson, Ruth D, Fiscutean, Andrada, Robles-Espinoza, Carla Daniela, De La Vega, Francisco M, et al., 2022. Virtual meetings promise to eliminate geographical and administrative barriers and increase accessibility, diversity and inclusivity. Nature Biotechnol. 40 (1), 133–137.

Xu, Ying, Wang, Dakuo, Collins, Penelope, Lee, Hyelim, Warschauer, Mark, 2021. Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. Comput. Educ. 161, 104059.

Xu, Ying, Warschauer, Mark, 2020. What are you talking to?: Understanding children's perceptions of conversational agents. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–13.

Yee, Nick, Bailenson, Jeremy, 2007. The Proteus effect: The effect of transformed self-representation on behavior. Hum. Commun. Res. 33 (3), 271–290. http://dx.doi.org/10.1111/j.1468-2958.2007.00299.x.

Zierau, Naim, Wambsganss, Thiemo, Janson, Andreas, Schöbel, Sofia, Leimeister, Jan Marco, 2020. The anatomy of user experience with conversational agents: A taxonomy and propositions of service clues. In: International Conference on Information Systems. ICIS.-Hyderabad, India.