

Under threat but engaged: Stereotype threat leads women to engage with female but not male partners in math

Katherine R. Thorson^{a,*}, Chad E. Forbes^b, Adam B. Magerman^b, Tessa V. West^{a,*}

^a Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, United States

^b Department of Psychological and Brain Sciences, University of Delaware, 111 Wolf Hall, Newark, DE 19716, United States

ARTICLE INFO

Keywords:

Student engagement
STEM diversity
Stereotype threat
Dyadic interaction
Physiological linkage

ABSTRACT

This research tests how experiencing stereotype threat before a dyadic interaction affects women's engagement with peers during a dyadic math task. In a pilot study ($N = 167$; $M_{\text{age}} = 20.1$ years), women who completed a manipulation of stereotype threat (a socially evaluative math task in front of male evaluators) experienced greater subjective threat than did men. In Studies 1A and 1B, math-identified female undergraduates completed the stereotype threat or control (doing math alone) manipulation and then completed a collaborative math task with another female or male student (who completed the control task). Sympathetic nervous system responses were collected to measure physiological linkage—the effect of participants' physiological states on their partners' subsequent physiological states—as an indicator of attention to the partner. We also measured the number of math-related questions participants asked their partners and task performance. In Study 1A (female-female dyads; $N = 104$; $M_{\text{age}} = 19.9$ years), threatened women asked more questions than controls did and became physiologically linked to their partners when those partners were speaking about math. Threatened women performed comparably to controls. In Study 1B (female-male dyads; $N = 140$; $M_{\text{age}} = 20.0$ years), threatened women did not ask more questions of their male partners than controls did, nor did they show physiological linkage to their male partners. Women performed worse than men did, regardless of condition. When working with a female, experiencing stereotype threat outside of a working interaction leads women to engage more; this effect does not occur when with a male.

1. Introduction

Despite efforts to improve women and girls' participation in Science, Technology, Engineering, and Math (STEM), the gender gap remains high across education and employment levels (National Science Foundation, 2017; U.S. Department of Commerce, 2013). Many psychological processes have been implicated in producing gender disparities in STEM, and the study of these processes has historically focused on people's experiences as they work and perform alone (e.g., Schmader, 2002; Schmader & Johns, 2003; Spencer, Steele, & Quinn, 1999). Recently, however, dyadic and group learning has been on the rise, and, thus, understanding what processes contribute to gender disparities in these interactions is imperative. Numerous studies of collaborative learning in STEM at the undergraduate level have shown enhanced exam and grade performance (e.g., in chemistry and biology; Drane, Smith, Light, Pinto, & Swarat, 2005; Tien, Roth, & Kampmeier, 2002), increased retention, better peer relationships, and improved attitudes toward STEM (e.g., attitudes towards mathematics and STEM

material in general; Drane et al., 2005; Springer, Stanne, & Donovan, 1999) when students are able to work together. The ability to work effectively with others is now a requirement for undergraduate accreditation in many STEM fields, such as engineering (Board, 2017), as jobs in science and engineering focus heavily on group work (Wuchty, Jones, & Uzzi, 2007).

Although promising, there are still many challenges involved in facilitating productive social interactions between peers at the undergraduate level, and educational psychologists have been debating how and when collaborative learning is most effective (for a review, see Micari, Pazos, Streitwieser, & Light, 2010). One critical process implicated in creating successful collaborative learning is *social engagement* (Capraro, Capraro, & Morgan, 2013; Ceci, Williams, & Barnett, 2009; Freeman et al., 2014; Jones, Howe, & Rua, 2000). Social engagement refers to students' interactions with others that are directly focused on STEM content and includes direct engagement with others (such as seeking help and asking clarification questions; Fredricks, Blumenfeld, & Paris, 2004; Stout, Dasgupta, Hunsinger, & McManus,

* Corresponding authors.

E-mail address: katherine.thorson@gmail.com (K.R. Thorson).

<https://doi.org/10.1016/j.cedpsych.2019.03.012>

2011), as well as attending to people who are useful sources of information. Attention includes processes such as listening carefully to a lecture given by a professor or a lesson by a teammate, looking at the blackboard during instruction, and listening to a peer respond to a question (Birch & Ladd, 1997; E.S. Shapiro, 2011).

The present research examines social engagement between university students who are highly identified with math while they work together in pairs to solve math problems. We specifically focus on how the experience of *stereotype threat* for women prior to working with an assigned female or male partner influences how engaged both partners are with each other during the dyadic task and how attentive they are to them. Utilizing a multi-method approach, we operationalize engagement in two ways. First, we measure the extent to which partners ask each other math-related questions during the task (Fredricks et al., 2004; Gasiewski, Eagan, Garcia, Hurtado, & Chang, 2012), which is an overt behavior. Two, we measure the extent to which partners show physiological linkage to each other on sympathetic nervous system (SNS) reactivity during the task. Linkage on SNS responses can indicate the extent to which individuals “track” the fluxes and flows of the intensity of their partners’ affective state and can provide online, moment-to-moment information regarding how attentive people are to their partners (see more details in Section 5; Kraus & Mendes, 2014; Thorson, West, & Mendes, 2018; West, Koslov, Page-Gould, Major, & Mendes, 2017). We also examine the differential effects of stereotype threat on math performance of the person who experienced stereotype threat and the partner of that person.

2. Social engagement in STEM interactions

As noted above, social engagement can facilitate collaborative learning in STEM. For undergraduate women, who are often less engaged with their peers in male-dominated STEM settings, such as in engineering and math (Dasgupta, Scircle, & Hunsinger, 2015; Grover, Ito, & Park, 2017), increasing engagement with peers may be a critical step in closing the gender gap. For example, undergraduate women who are engaged with their peers in engineering experience greater feelings of social belonging, which in turn buffer them from failures and setbacks associated with STEM drop-out (Dennehy & Dasgupta, 2017; Solder, Rowan-Kenyon, Inkelas, Garvey, & Robbins, 2012; Walton, Logel, Peach, Spencer, & Zanna, 2015).

To date, however, there is little empirical work that directly examines social engagement in STEM by assessing behaviors in interpersonal working contexts as they naturally unfold over time (see Appleton, Christenson, & Furlong, 2008; Fredricks et al., 2004; Gasiewski et al., 2012 for theoretical reviews). The only research in this domain of which we are aware assesses how trained undergraduate female experts contribute to a group math task (e.g., whether they correctly explain how to compute the answer to a math problem) as a function of the group’s gender composition (Grover et al., 2017). Here, we focus on social engagement between peers, as peers can be either barriers or bridges to success in STEM settings. In the best case, peers help each other problem-solve and fulfill social inclusion goals, and in the worst case, peers distract or hinder problem-solving and have the power to socially isolate people.

3. The effect of stereotype threat on social engagement in STEM interactions

How might the experience of stereotype threat affect interactions between peers during a dyadic math task? For undergraduate women who are motivated to perform well in STEM fields, such as math, stereotype threat occurs when they are concerned that they will confirm the negative stereotype that women are not as competent as men within STEM (Logel, Peach, & Spencer, 2012; Spencer et al., 1999). Scholars have theorized that there are several important components to a stereotype-threatening experience, which we briefly review here. One,

stereotype threat is relevant to members of groups who are aware that there is an existing stereotype that members of their group are not as competent as members of other groups in a particular domain (e.g., women are not as competent as men on math tasks; Schmader, Johns, & Forbes, 2008; Spencer et al., 1999). Two, members from the negatively stereotyped group need to perform on a task that is relevant to the stereotype in a context where they are being evaluated (e.g., they need to perform math in front of evaluators). Three, the stereotype needs to be “salient” to group members when they are performing, which can be accomplished in a number of ways, including being directly evaluated by members of the group who are stereotyped as performing better than members of one’s own group (in this case men), or, as some scholars have shown, simply performing alongside members of the other group. For example, undergraduate women can experience stereotype threat when they are outnumbered by men when solving math problems (Inzlicht & Ben-Zeev, 2000) or when they take a test that measures quantitative capacity and are told that men outperform women (Schmader & Johns, 2003; Spencer et al., 1999).

Within stereotype threat research, the focus has historically been on how manipulations of threat affect performance (see recent meta-analyses by Flore & Wicherts, 2015; Nguyen & Ryan, 2008; Pennington, Heim, Levy, & Larkin, 2016), and performance differences (or lack thereof) have been used as evidence that stereotype threat occurred (e.g., Spencer et al., 1999). However, it is important to distinguish between the psychological experience of stereotype threat—which often manifest as physiological and psychological threat (e.g., feeling like the demands of the situation outweigh one’s resources)—from the outcomes of this experience (Schmader, Forbes, Zhang, & Mendes, 2009). Beyond performance, numerous studies have examined the impact of stereotype threat on women’s interest in math at the undergraduate level (e.g., Davies, Spencer, Quinn, & Gerhardstein, 2002; Murphy, Steele, & Gross, 2007; Shapiro, Williams, & Hambarchyan, 2013). Critically, scholars have theorized that the chronic experience of stereotype threat may lead women to opt out of STEM fields—for example, to major in English over physics or to work as a science journalist instead of a lab technician (Steele, Spencer, & Aronson, 2002). Thus, stereotype threat might also lead women to be less engaged with their peers during a dyadic math task—a possibility we examine in the present research.

In considering how stereotype threat regarding math will affect engagement when women are working with others, research on stress and coping within dyadic interactions provides a useful framework for forming hypotheses. According to a stress and coping approach, not all experiences of stress are created equal; stress can either be appraised as *threat* (i.e., people feel like their demands outweigh their resources) or *challenge* (i.e., people feel like their resources match or outweigh their demands; Blascovich & Mendes, 2010; Lazarus & Folkman, 1991). Accordingly, stressors associated with the experience of stereotype threat can be experienced as threats or as challenges (Pennington et al., 2016). For example, one study showed that first-year female undergraduates in engineering who were outnumbered by men were more likely to report feeling greater threat relative to challenge (Dasgupta et al., 2015). For women in math, when stressors are experienced as threats, they lead to poorer psychological and performance outcomes than when they are experienced as challenges (Ben-Zeev, Fein, & Inzlicht, 2005; Johns, Inzlicht, & Schmader, 2008; Vick, Seery, Blascovich, & Weisbuch, 2008). For example, among female undergraduates who showed high SNS arousal while taking a practice version of the math Graduate Record Examination, those who were able to appraise their physiological state as a challenge scored higher on the GRE than those who did not (Schmader et al., 2009; see also Jamieson, Nock, & Mendes, 2012). Relevant to the present research, how people appraise stressors can affect how they behave towards their partners, and specifically, how much they engage with their partners. Whether people experience a stressor as a challenge or threat can affect whether they engage with their partners or “freeze” (Blascovich & Mendes, 2010; Trawalter,

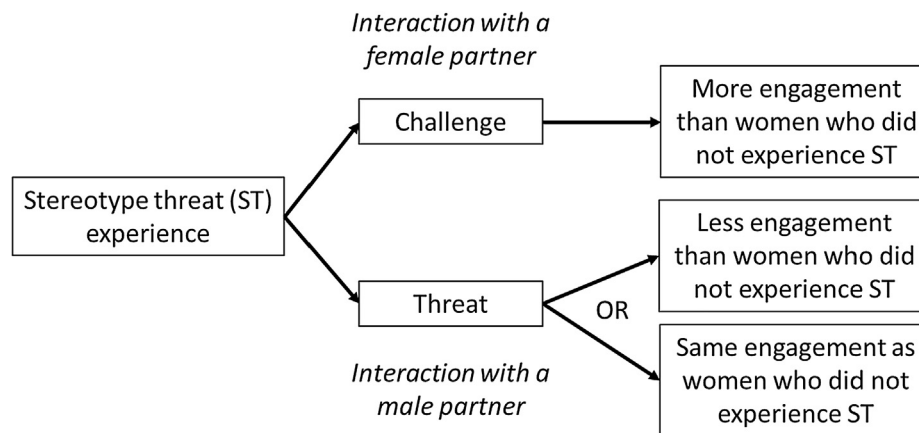


Fig. 1. Model for how stereotype threat might affect women's engagement when interacting with a female partner or a male partner.

Richeson, & Shelton, 2009). If they appraise stressors as a challenge, they are likely to engage fully with their partners; in the present research, this would mean asking their partners more questions and attending closely to them. However, if they appraise the psychological stressor as a threat, they are likely to freeze with their partners—failing to engage at all.

Following a stress and coping framework, we propose that women will experience a stereotype threat manipulation regarding their gender and math as subjectively threatening (which we demonstrate in the pilot study), but they may not continue to experience threat *during* the dyadic interaction. As such, the degree to which they engage with their partners will vary depending on the gender of their partner. In the next section, we utilize a stress and coping framework to theorize about how women who have experienced stereotype threat will engage with female and male working partners.

3.1. Stereotype threat during female-female interactions

In Fig. 1, we present a guide of how we propose stereotype threat will affect interactions that women have with female and male partners. We propose that following a threat experience in which women are evaluated by men while performing math, in female-female pairs, women will demonstrate a behavioral pattern consistent with a challenge state, engaging more with their female partners relative to those who have not undergone stereotype threat. We base this hypothesis on research showing that (a) women at the undergraduate level often feel more comfortable engaging with female peers than male peers, even in socially evaluative settings within STEM fields, such as engineering and math (Dasgupta et al., 2015; Grover et al., 2017). This suggests that female partners do not invoke a threat response in women and (b) the presence of other females in undergraduate STEM fields, such as engineering and math, can buffer against stereotype threat effects, as women serve to protect threatened women from the aversive effects of stereotype threat on perceptions of one's own math ability and feelings of self-efficacy (Dennehy & Dasgupta, 2017; Inzlicht & Ben-Zeev, 2000; Marx & Roman, 2002). To the extent that women who experience stereotype threat are able to “harness” their arousal during the dyadic interaction to their benefit, then they may show even stronger engagement with their partners than women who do not experience stereotype threat.

3.2. Stereotype threat during female-male interactions

In the present research, we also examine whether male interaction partners serve as a stressor during the interaction for women. Although the presence of men can evoke stereotype threat effects (Dennehy & Dasgupta, 2017; Inzlicht & Ben-Zeev, 2000; Sekaquaptewa &

Thompson, 2003), it remains to be seen whether male peer partners heighten existing stereotype threat experiences for women after they have undergone a threat manipulation. We theorize that if male partners do serve as stressors, there are two possible outcomes for women who interact with men.

The first possible outcome, as illustrated in Fig. 1, is that women who have undergone a threat manipulation will be less engaged with their male partners than women who have not; even though they are not the source of the threat themselves, male partners may operate as a “chronic reminder” of the stereotype threat experience and exacerbate the effects women experienced going into the encounter. Indeed, prior research has demonstrated that individuals can carry over stress from one interaction to another (Waters, West, & Mendes, 2014), particularly when people are interacting under conditions that are also stressful (Liu, Rovine, Cousino Klein, & Almeida, 2013; Saxbe & Repetti, 2010). If this is the case, then within male-female dyads, we would expect women who have undergone the threat experience prior to the interaction to be less engaged (ask fewer questions; exhibit weaker physiological linkage) than those who have not undergone the threat experience (tested in Study 1B). In addition, we would also expect that women under threat would be less engaged with male partners than women under threat would be with female partners because male partners are exacerbating the threat experience (a possibility we test for in the cross-study comparisons).

The second possibility is that male interaction partners can evoke a threat response for all women. To this end, having a male partner may “level the playing field” between women who entered the dyadic encounter under threat and those who did not. Indeed, women subjectively feel and appear less engaged when interacting in STEM contexts with men, regardless of level of acquaintance (e.g., in academic STEM fields and in undergraduate and professional engineering), than with women even without explicit manipulations of stereotype threat (Dasgupta et al., 2015; Hall, Schmader, & Croft, 2015; Holleran, Whitehead, Schmader, & Mehl, 2011). These results suggest that the mere presence of a male partner is sufficient to invoke the psychological experience of stereotype threat.

If this is the case, then within male-female dyads, we would expect women who have undergone the threat experience prior to the interaction to be equally engaged with their partners (i.e., exhibit similar physiological linkage, ask similar amount of questions) as those who have not undergone the threat experience (tested in Study 1B). In addition, if men lead to threat for all women, then we would also expect all women paired with men to be less engaged with their male partners than women paired with women. We test this possibility in the cross-study comparisons.

We note that in both of these cases, women will be less engaged with their male partners than their female partners, but in the former,

stereotype threatened women who are paired with men are the least engaged among all types of women because men exacerbate the threat experience.

4. The effect of stereotype threat on performance in STEM interactions

Lastly, we examine the question: How does the experience of stereotype threat affect performance in STEM interactions? As noted above, historically, the predominant outcome in studies on stereotype threat has been performance. Recent attempts to replicate the effect of stereotype threat on math performance when working alone have been mixed; some research has found no effects of stereotype threat on women's performance (relative to women who did not experience stereotype threat; among adults: Finnigan & Corker, 2016; among 4th through 12th graders: Ganley et al., 2013; among undergraduates: Gibson, Losee, & Vitiello, 2014), and some has found negative effects (i.e., stereotype threat hinders performance; among undergraduates: Schmader, 2002; Schmader & Johns, 2003; Spencer et al., 1999). In addition, a recent meta-analysis that focuses on the effects of stereotype threat for children on math performance found some evidence of an effect (Flore & Wicherts, 2015), which they interpret with caution given evidence of a strong file drawer effect. A second recent meta-analysis, which focuses on performance for adults, finds a mixed pattern of results and strong evidence that performance differences depend on contextual factors (Pennington et al., 2016).

These findings are consistent with literature indicating that there are several potential moderating factors that explain when and how stereotype threat affects performance (such as how much women identify with math and whether they are aware of the gender-math stereotype; Gibson et al., 2014; Nguyen & Ryan, 2008). Among these, the psychological appraisal of one's resources compared to one's demands—in other words, whether one is threatened (as opposed to challenged)—is critical. Threat can deplete cognitive resources, such as working memory capacity, which can then impair performance (Johns et al., 2008; Schmader & Johns, 2003). Thus, similar to the effects of challenge and threat on social engagement, challenge and threat can also affect performance, such that greater threat leads to worse performance (Ben-Zeev et al., 2005; Jamieson, Mendes, Blackstock, & Schmader, 2010; Schmader et al., 2009).

In the present research, we explore whether women who work with women will demonstrate better performance, as well as engage more with their partner, when under threat. In addition, we explore whether women who work with men perform worse when under threat than those who are not under threat, which, if found, would be consistent with classic findings of stereotype threat (Inzlicht & Ben-Zeev, 2000; Spencer et al., 1999).

To test the idea that men serve to elicit stereotype threat in women, we not only compare the performance of men to women; but also, we compare control condition women who have interacted with men to control condition women who have interacted with women. This latter comparison is critical for testing the idea that male peers serve as stressors for women, leading them to perform worse than women who have not interacted with male partners.

Finally, we examined changes over time in performance. We did not have specific hypotheses about trajectories in performance over time as a function of gender and stereotype threat. However, prior work on threat and stress has shown that the effects of stressors generally weaken over time, with people adapting and returning to physiological baselines over time (McEwen, 2007; Selye, 1946). Therefore, any effects we observe on performance may be the strongest at the beginning of the interaction, when the stereotype threat manipulation is closest in time to the dyadic interaction.

5. Research overview

The present studies are the first to our knowledge to provide a comprehensive investigation of how experiencing stereotype threat prior to a novel interaction (and without one's interaction partner knowing about the stereotype threat manipulation) influences two processes related to engagement and performance. In both partners—undergraduate students highly identified with math—we measure behaviors and physiology over the course of the interaction to directly capture social engagement as it naturally unfolds. To manipulate stereotype threat, women who were highly identified with math performed a math test, framed as diagnostic of math intelligence, in front of male evaluators. We reasoned that these types of threatening experiences—being evaluated by others on one's domain-specific aptitude—are particularly common in STEM (e.g., students giving class presentations, and students performing lab technique in front of others). Furthermore, these experiences are critical for understanding social engagement because they are often followed by learning- and work-based interactions with others (Capraro et al., 2013; Wuchty et al., 2007). Moreover, we argue that it is important to capture the psychological and physiological experience of threat *prior* to measuring key outcomes of interest. To ensure that the manipulation induces threat, we apply well-replicated research on how social evaluation can evoke the psychological experience of threat (Dickerson & Kemeny, 2004) by measuring a combination of physiological and self-reported measures immediately following the stereotype threat manipulation (Pilot Study). Our goal is to demonstrate that the threat manipulation uniquely leads women to subjectively interpret their experience as threat. Only after establishing this do we then test *how* the manipulation affects our outcomes of interest (Studies 1A and 1B).

In Study 1A, we test whether women engage more with women when they are under stereotype threat (versus not) by comparing female-female dyads in which one woman experienced stereotype threat and the other did not to dyads in which neither woman experienced stereotype threat. In Study 1B, we tested whether women who experienced stereotype threat prior to an interaction with a male partner would be less or similarly engaged compared to women who did not experience stereotype threat prior to an interaction with a male partner. Given the critical importance of performance outcomes in educational domains, we also examined the performance of women who experienced stereotype threat (relative to those who did not experience stereotype threat) in both studies.

5.1. Measuring social engagement

We measured women's social engagement while working on a math task with a female or male peer in two ways—(1) by directly assessing an overt behavior and (2) by assessing a dyadic, physiological measure of behavioral engagement that represents a more automatic process.

5.1.1. Question asking about math

First, we measured the number of questions women asked their partners about the math task. Asking questions provides opportunities to gain information and certainty and is a direct measure of social engagement within learning and performance settings (Fredricks et al., 2004; Gasiewski et al., 2012). Because we were interested directly in STEM-relevant social engagement, we measured not just the number of questions that partners asked each other during the math task (which might also be about the classes they were taking or where they were living, for example) but specifically the number of questions asked about the math task.

5.1.2. Physiological linkage

As a second measure of social engagement, we measured the extent to which women became physiologically linked to their partners during the math task—that is, how much one woman's physiological state was

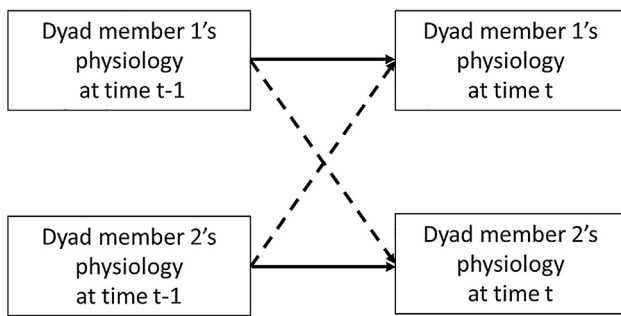


Fig. 2. Model of physiological linkage. The solid lines represent stability or autoregressive paths, where a dyad member's physiology at one time point predicts their own physiology at a later time point. The dashed lines represent the influence or cross-lagged paths, where a dyad member's physiology at one time point predicts the other dyad member's physiology at a later time point. When dyad member 2's physiology at time t-1 predicts dyad member 1's physiology at time t, dyad member 1 is said to be physiologically linked to dyad member 2.

predicted by her partner's physiological state at a prior time point (see Fig. 2). Physiological linkage between partners can be used as a dynamic measure of the degree to which partners are attuned to each other during interactions (Kraus & Mendes, 2014; Palumbo et al., 2016; Thorson et al., 2018; West et al., 2017) and is related to interpersonal accuracy (Levenson & Ruef, 1992).

Using linkage to understand attention has several benefits: linkage can be captured within naturalistic learning contexts, it does not rely on outside observers to make subjective judgments about how attentive students appear to be (e.g., E.S. Shapiro, 2011), and it can be assessed dynamically throughout an interaction. In this study, we measured physiological linkage of students' sympathetic nervous system (SNS) responses, assessed with non-invasive cardiovascular measures. Sympathetic nervous system activity can be interpreted broadly as a measure of affective intensity (Mendes, 2016), and thus, linkage on SNS responses indicates the extent to which individuals "track" the fluxes and flows of the intensity of their partners' affective state. Psychologically, linkage occurs when the physiological response of one dyad member is associated with signals that the other dyad member notices; the second dyad member "picks up on" these cues and then experiences a similar physiological state (Thorson et al., 2018).

In Thorson et al. (2018), we theorize that linkage is conditional—people are most strongly linked to partners when those partners are engaging in behaviors that are motivationally relevant. For example, in cross-race interactions between African Americans and European Americans, African Americans show physiological linkage to European American partners, when those partners are leaking cues of anxiety (e.g., appear tense and uncomfortable)—behaviors that are associated with being prejudiced (West et al., 2017). For African Americans, detecting prejudice in a partner is motivationally relevant (Dovidio, 2001). In Kraus and Mendes (2014), lower-status individuals show linkage to higher status people who have influence over their outcomes.

Measures of physiological linkage can also provide insight into when people are paying the most attention to their partners because linkage tends to be stronger when people's partners engage in behaviors that they are motivated to care about (Thorson, Dumitru, Mendes, & West, in preparation; West et al., 2017). Thus, to study STEM-relevant social engagement, we examined whether women were particularly attentive to their peers when those peers were engaging in motivationally-relevant social behaviors—that is, behaviors related to the task of solving math problems together. One motivationally-relevant social behavior in this working STEM context is actively talking about the math problems (as opposed to looking at one's computer screen and working in silence). Therefore, as a second measure of STEM-relevant social engagement, we tested whether female students were more

physiologically linked to partners when those partners were talking about math.

6. Pilot study

The present threat manipulation involves social evaluation, which can elicit robust responses for both men and women (Dickerson & Kemeny, 2004). Thus, we first sought to demonstrate that only women interpreted stress during the stereotype threat manipulation as threat. Stereotype threat for women in STEM occurs when women are concerned they will confirm the gender-STEM stereotype that women are not as competent as men within STEM (Spencer et al., 1999). Therefore, to manipulate experiences of stereotype threat, women must be concerned about their STEM aptitude (as opposed to aptitude in another field, for example) and concerned about their gender—in other words, their gender needs to be salient to them. To manipulate concern about STEM aptitude, our manipulation of stereotype threat involved a task that was framed as diagnostic of students' math intelligence. To manipulate concern about their gender, the task was completed in front of male evaluators, with a male experimenter (see Forbes, Duran, Leitner, & Magerman, 2015; Johns et al., 2008; Sekaquaptewa & Thompson, 2003 for manipulations of math stereotype threat among female undergraduates with a male experimenter). This task framing and the presence of males have been used both separately and in combination to elicit math stereotype threat effects among female undergraduates in past research (Grover et al., 2017; Inzlicht & Ben-Zeev, 2000; Marx & Roman, 2002). Recently, scholars have argued that both components are needed (task framing and the presence of males), as both pieces fully capture the theoretical basis for math stereotype threat effects (Forbes et al., 2015; Johns et al., 2008; Sekaquaptewa & Thompson, 2003).

Participants were asked to complete a diagnostic mental math test in front of two male evaluators using a 2 (gender: male or female) by 2 (threat condition: threat or control) between-subjects design. We measured self-reported perceptions of task demands and resources, where a psychological state of "threat" is characterized as perceiving fewer resources than demands. We based this approach on theorizing by Blascovich and Mendes (2010) indicating that neither resources nor demands alone can indicate where people fall on the challenge vs. threat continuum because threat occurs when the appraisal of demands is greater than the appraisal of resources. For example, two people could both perceive a task as highly demanding, but if one person perceives having greater resources, he/she will experience psychological challenge. If the other person does not perceive greater resources than demands, he/she will experience threat. Thus, the "demands" perceptions might be the same for the two people, but the experience of challenge vs. threat is quite different. To the extent that our stereotype threat (ST) manipulation evokes gender-specific threat, we hypothesized that only women would report greater demands relative to controls; men would not report greater demands relative to controls. Moreover, women in the threat condition would report greater demands relative to men in the threat condition.

6.1. Method

Additional methodological and analytic details for all three studies are provided in the Supplemental Materials (SM); a video of the procedure is provided at https://youtu.be/_glakg0D85Q; measures, datasets, and syntax are available on the Open Science Framework (OSF) at <https://osf.io/gpw4j/>. All three studies received research ethics committee approval.

6.1.1. Design and participants

Participants were math-identified college students who had knowledge of the stereotype that men are better at math than women (see Section 6.1.3 below for how we determined this; 167 participants;

69 Asian or Pacific Islander, 43 White, 7 Black, 22 Hispanic, 22 multiracial, and 4 “other”; $M_{age} = 20.12$, $SD_{age} = 1.30$; n_s per cell: 41 ST females, 42 control females, 42 ST males, 42 control males); seven participants who are not included in the numbers above were excluded *a priori* (three because of experimenter error and four because of participant non-compliance). Given the focus on stereotypes and social categories in this work, we recognize that the terms women/men might be more appropriate than female/male. However, for ease of clarity, we often use the terms female/male as adjectives that describe other nouns (e.g., “female partner” or “male control”). Because these conditions can affect cardiovascular responding, participants in all three studies reported in this paper (Pilot Study, Study 1A, and Study 1B) were pre-screened to ensure that they had a body mass index lower than 30, were not taking cardiac medications, were not pregnant, and did not have a pacemaker or a doctor’s diagnosis of a heart arrhythmia or hypertension (Blascovich, Vanman, Mendes, & Dickerson, 2011). Because we did not wish to put already vulnerable populations through the stress of the stereotype threat manipulation, we also screened participants to ensure they did not have a history or diagnosis of a psychiatric illness. Participants in all three studies in this paper completed the research in exchange for partial course credit, \$15, or \$20 (depending on semester) and attended the same university in the Northeastern region of the United States. None of the participants in any of the three studies was allowed to participate in more than one of the studies (i.e., there was no overlap in participants across the three studies).

6.1.2. Procedure

After arrival at the lab, participants in the ST condition were instructed to count backwards from 2023 to zero in 17-step sequences as quickly and correctly as possible (similar to the Trier Social Stress Test; Kirschbaum, Pirke, & Hellhammer, 1993). The task was framed as an experimental math test diagnostic of math intelligence. Two White male evaluators were present and monitored participants for mistakes and instructed participants to start over after a mistake. In the control condition, participants completed the same task, framed as an experimental problem-solving exercise, on the computer with no evaluators present. The computer forced them to start over after a mistake. Both the ST and control tasks lasted for five minutes. After the manipulation, participants were told they would be completing either a standardized, diagnostic math test (ST condition) or a standardized problem solving exercise (control condition) with a partner.¹ They then completed a measure of demands and resources regarding the upcoming test/task. In all studies, participants who completed the ST manipulation had a White male experimenter; participants who completed the control manipulation had a White female experimenter.

6.1.3. Measures

6.1.3.1. Knowledge of gender-math stereotype. Based on prior math stereotype threat research (Aronson et al., 1999; Johns et al., 2008; Spencer et al., 1999), we pre-screened participants to ensure that each participant had knowledge of the stereotype that men are better at math than women (i.e., that each participant responded 3 or lower to the following question: “Regardless of what you think, what is the stereotype that people have about women and men’s math ability, in general?” where 1 = *men are much better than women*, 4 = *men and women are the same*, and 7 = *women are much better than men*; $M = 1.84$, $SD = 0.68$).

6.1.3.2. Math identification. Prior stereotype threat work (Aronson

¹ We also manipulated and revealed the gender of the presumed upcoming interaction partner prior to participants completing the resource/demands index. We did not include partner gender as a factor in the analyses for the resource/demands index as it did not affect the direction or significance of the results.

et al., 1999; Keller, 2007; Steele, 1997) has demonstrated that people highly identified with a domain are the most susceptible to stereotype threat effects within that domain. Therefore, we pre-screened participants to ensure that they identified with math by scoring an average of 5 or higher on a 9-question measure assessing identification with math ($\alpha = 0.72$; $M = 5.81$, $SD = 0.57$; borrowing items from Crocker, Luhtanen, Cooper, & Bouvrette, 2003; Forbes & Schmader, 2010; and Major & Schmader, 1998). All questions were answered on a 1 (*strongly disagree*) to 7 (*strongly agree*) scale, and an example item is as follows: “It is important for me to be good at tasks that require the use of math.”

6.1.3.3. Appraisals of demands and resources. We measured perceived demands (6 items, e.g., “I think this task/test represents a threat to me,” $\alpha = 0.73$) and resources (5 items, e.g., “I view this task/test as a positive challenge,” $\alpha = 0.74$; see Jamieson, Peters, Greenwood, & Altose, 2016). Consistent with prior research (Mendes, Gray, Mendoza-Denton, Major, & Epel, 2007), we averaged items across the demands and resources subscales and computed a demand/resource ratio. Higher values indicate greater demands relative to resources (i.e., a psychological state of “threat”).

6.2. Results

For all analyses in the paper, we use an alpha of 0.05 to determine statistical significance (Frick, 1996; Knapp, 2015). To control for Type I error in post-hoc pairwise comparisons, we have applied a Bonferroni correction (Abdi, 2007). Applying a Bonferroni correction is a common approach to controlling for Type I errors. To do this, we took the *p*-values obtained from each pairwise comparison and multiplied each one by the number of comparisons that were done. As is convention, we then reported this adjusted *p*-value and compared it to an alpha of 0.05 to determine significance.

6.2.1. Demand/resource index

A main effect of condition, $F(1, 155) = 12.16$, $p = .001$, $\eta_p^2 = 0.073$, indicates that participants in the ST condition perceived greater demands relative to resources than those in the control condition perceived. There was no main effect of gender, $F(1, 155) = 2.89$, $p = .091$, $\eta_p^2 = 0.018$, but a significant Condition \times Gender interaction, $F(1, 155) = 3.92$, $p = .049$, $\eta_p^2 = 0.025$. As shown in Table 1, women in the ST condition experienced greater demands (relative to resources) than did men in the ST condition ($p = .010$), but there was no difference between males and females in the control condition ($p = .84$). Women in the ST condition also experienced greater demands relative to resources than did those in the control condition ($p < .001$). For men, there was no effect of condition ($p = .28$). Although the effect size for the Condition \times Gender interaction is small, these findings nevertheless provide evidence that the manipulation evoked threat for females but not males.²

² Upon the request of reviewers, we estimated a multivariate model in which we predicted the demand and resource appraisals from gender, condition, type of appraisal (demand vs. resource) and all two- and three-way interactions. We found a marginal Type \times Gender \times Condition interaction, indicating that there was only weak evidence that the Gender \times Condition varied as a function of the type of appraisal, $F(1, 156) = 3.74$, $p = .055$. Follow-up analyses showed that the Gender \times Condition interaction was marginally significant for demands, $t(274.91) = -1.81$, $p = .071$, and nonsignificant for resources, $t(274.91) = 1.38$, $p = .17$. The means for both scales followed similar patterns (the greatest demands were shown by women in the threat condition and the fewest resources were shown by women in the threat condition). Thus, these results do not provide strong evidence that the demand vs. resource appraisals were affected differently by participant’s gender and condition.

Table 1
Pilot study: Descriptive statistics for the demand/resources index.

| | ST condition | | Control condition | |
|---------------------|--------------|-------|-------------------|-------|
| | Females | Males | Females | Males |
| Mean | 0.97 | 0.80 | 0.71 | 0.72 |
| Standard deviation | 0.36 | 0.38 | 0.20 | 0.21 |
| 95% CI: Lower bound | 0.87 | 0.70 | 0.62 | 0.63 |
| 95% CI: Upper bound | 1.07 | 0.89 | 0.80 | 0.82 |

Note. Higher values indicate greater demands relative to resources. CI = confidence interval.

7. Study 1A

In Study 1A ($n_{dyads} = 52$, $n_{participants} = 104$), we examined how stereotype threat, experienced prior to a dyadic interaction, affects how STEM-identified women work with other STEM-identified women.

7.1. Method

7.1.1. Design

Dyads either had one woman who received the ST manipulation (“ST targets”; $n = 35$) and one participant who received the control manipulation (“ST non-targets”; $n = 35$; see Fig. 3), which we call “stereotype threat dyads”, or two women who both received the control manipulation (“controls”; $n = 34$), which we call “control dyads”. A video of the procedure is provided at <https://youtu.be/mFakTnx5vuY>.

7.1.2. Participants

Participants were math-identified college students who had knowledge of the stereotype that men are better at math than women (104 participants; 49 Asian or Pacific Islander, 32 White, 8 Black, 6 Hispanic, 5 multiracial, and 4 “other”; $M_{age} = 19.91$, $SD_{age} = 1.21$); nine dyads who are not included in the numbers above were excluded *a priori* (seven for experimenter error and two for participant non-compliance). Dyads were not matched on race.

We conducted power analyses for the ability to detect physiological linkage within dyadic interactions, given that linkage effects tend to be very small, and thus, this method provides a conservative power estimate (based on our prior research; e.g., West et al., 2017). To do so, we utilized a simulation method that is illustrated in Bolger, Stadler, and Laurenceau (2011), Lane and Hennes (2018) and Thorson et al. (2018). We first provided estimates for a typical range of linkage estimates and used these estimates to simulate data for 1000 hypothetical studies with 50–70 dyads each. Once the data were simulated, we analyzed each of the 1000 samples individually and output the number of times the effect of interest was significant (at the $p < .05$ threshold). Across all analyses, we had between 50 and 94% statistical power to detect physiological linkage for sample sizes ranging from 50 to 70 dyads

(Study 1A has 52 dyads and Study 1B has 70 dyads). Given these sample sizes, we had between 75 and 99% statistical power to detect a small to medium effect of the stereotype threat manipulation on performance.

7.1.3. Procedure

Participants arrived separately and recorded a 5-minute physiological baseline in separate rooms (see Fig. 3). Next, ST targets completed the ST task, and ST non-targets and controls completed the control task (both same as the pilot study). Immediately after the ST or control task was completed, dyad members were moved to the same room. As soon as they were moved, the experimenters provided instructions for completing the next task, where dyad members worked to solve 27 math problems. In total, about five minutes passed from the time the ST/control task was completed and the dyadic math task began. The problems were framed as a standardized, diagnostic math test for ST targets, in keeping with the intelligence-diagnostic framing of the ST manipulation, and as a standardized problem solving exercise for ST non-targets and controls, in keeping with the non-diagnostic framing of the control manipulation. The math problems were presented on a computer screen, and participants were given 30 s to answer each question while working alone, 30 s to discuss the problem with their partner, and 5 s to provide a final answer. They were informed of these time limits before the task began; however, there was no timer on the computer screen. Participants rotated between easy, medium, and hard questions. If participants did not respond within the allotted time, the computer automatically moved on to the next question, and the item was marked as unanswered (treated as incorrect). Participants had to solve the problems mentally; they were not allowed pencil, paper, or calculators to help solve the problems.

7.1.4. Measures

7.1.4.1. Knowledge of gender-math stereotype. We used the same item as in the pilot study to ensure that all participants were aware of the stereotype that men are better at math than women ($M = 1.69$, $SD = 0.74$).

7.1.4.2. Math identification. We used the same measure as in the pilot study to ensure that participants were highly identified with math ($\alpha = 0.73$; $M = 5.79$, $SD = 0.57$).

7.1.4.3. Physiological measures. We measured sympathetic nervous system activity (SNS; the branch of the autonomic nervous system which mobilizes the body for action; Fox, 2006) via pre-ejection period (PEP; one of the purest measures of SNS; Schachinger, Weinbacher, Kiss, Ritz, & Langewitz, 2001), which is the amount of time during a cardiac cycle between the left ventricle of the heart contracting and the aortic valve opening. PEP is associated with momentary changes in intensity of affective states and is responsive to changes in a short time frame (Mendes, 2016). We employed electrocardiography (ECG) and impedance cardiography (ICG) to obtain measurements of PEP. We

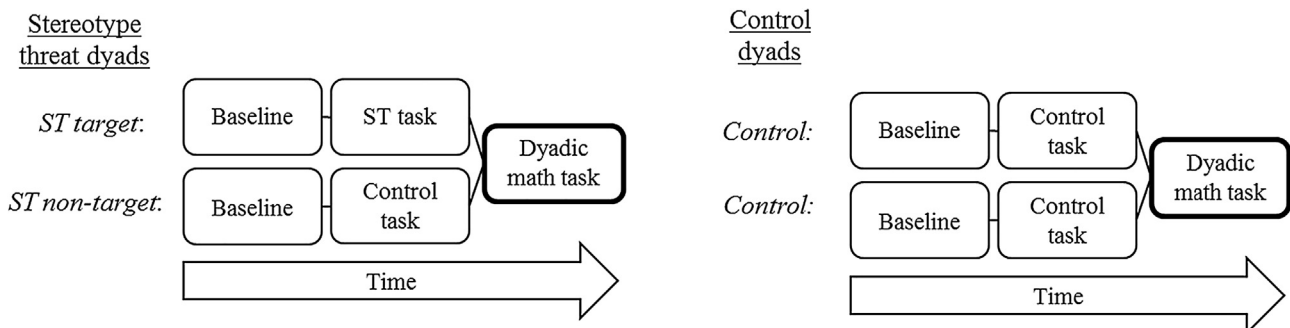


Fig. 3. Overview of the procedures in Studies 1a and 1b for each dyad member in both types of dyads (stereotype threat and control). Bold outlines indicate that dyad members were together; at all other times, dyad members were in separate rooms.

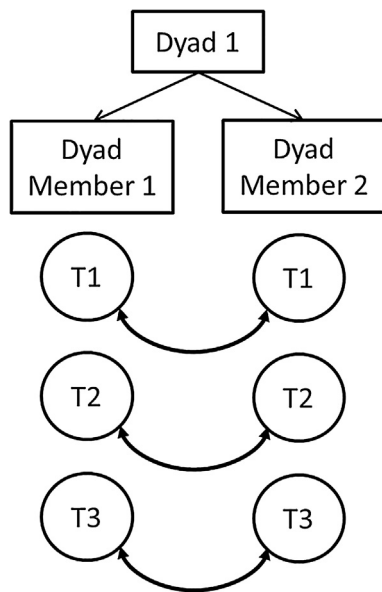


Fig. 4. A two-level crossed model where each dyad member has provided physiological 'data at the same time points.

recorded ICG and ECG responses using an integrated system (Biopac MP150, Biopac Systems, Goleta, CA) with amplifiers for ECG (ECG100C) and ICG (NICO100C). We used band electrodes in a standard tetrapolar configuration for the recording of ICG responses, and two snap electrodes in a modified Lead II configuration (near the right clavicle, below the ribcage on the left side of the torso) for the recording of ECG responses. A 400 μ A current was passed through the outer band electrodes, and Z0 and its first derivative, $\Delta z/\Delta t$, were recorded from the inner bands.

After the study, physiological data were analyzed in 30-s intervals using Mindware's impedance cardiography software (IMP 3.0.25, Mindware Technologies, Gahanna, OH), and PEP measurements were calculated as the amount of time between the Q point on the ECG wave (when the left ventricle contracts) and the B point on the $\Delta z/\Delta t$ wave (when the aortic valve opens). We visually inspected all intervals and manually selected the Q and B points when they were incorrectly identified by the software. Several papers have been written about how to detect the Q and B points, as well as the pros and cons of different detection methods (see [Berntson, Lozano, Chen, & Cacioppo, 2004](#); [Lozano et al., 2007](#); [Sherwood et al., 1990](#)); however, manual detection of the points by trained researchers is considered to be the most accurate method for Q and B point detection ([Blascovich et al., 2011](#)). We selected the B point as the notch at the beginning of the longest upstroke before the Z point ([Lozano et al., 2007](#)). We computed reactivity scores by subtracting baseline PEP responses (the last 30-s interval of baseline) from PEP responses in 30-s intervals throughout the dyadic task (see [Waters, West, Karnilowicz, & Mendes, 2017](#), & [West et al., 2017](#) for the same procedures).

7.1.4.4. Math questions involved the addition of fractions, multiplication of whole numbers, and division of whole numbers. The math questions were drawn from a pilot test of questions that ranged in difficulty. In the pilot test, which was completed by undergraduates, easy questions were solved accurately 78.70% of the time, medium questions 49.27% of the time, and hard questions 38.86% of the time (the pilot test of these items is reported in [Forbes, Amey, Magerman, Duran, and Liu \(2018\)](#), and the items are listed on this paper's OSF page).

7.1.4.5. Questions asked and talk time. Videos of participants during the dyadic math task were coded by trained research assistants blind to the study hypotheses. After training, one coder overlapped with each

additional coder for 10% of the videos. Interrater reliability was assessed using a one-way random effects single-measures ICC ([McGraw & Wong, 1996](#)). The resulting ICCs were in the excellent range (ICC for questions asked = 0.75; ICC for talk time = 0.90; [Cicchetti, 1994](#)) indicating that questions asked and talk time were coded similarly across coders. The number of questions participants asked that were specifically related to the math task and the amount of time participants spent talking about math (i.e., talk time in the 30-s intervals they were exclusively talking about math) were each summed across 3-question (easy, medium, hard) segments, resulting in 9 data points per participant ([Heyman, Lorber, Eddy, & West, 2013](#)). Values for talk time represent percentages. We recorded each dyad member in Studies 1A and 1B using separate video cameras.

7.1.5. Analytic strategy for dyadic analyses

Because data are dyadic and measured over time, we estimated two-level crossed models to account for non-independence in participants' responses. Dyad members are nested within dyads, and because the level of repeated measure is the same for both members of the dyad, repeated measure and dyad member are crossed (not nested; see this paper's OSF page for all syntax; also see Chapter 13 of [Kenny, Kashy, & Cook, 2006](#), and [West, 2013](#), for more details on two-level crossed models). Two-level crossed models are appropriate because dyad members have provided physiological data at the same time points so one can estimate the correlation of errors within each time point for the two dyad members (see [Fig. 4](#)). For example, person A's time 1 error is correlated with person B's time 1 error.

Given that partners in the control dyads cannot be distinguished from one another based on a meaningful theoretical factor that is dichotomous (i.e., they are both in the control condition), we treated all dyads as indistinguishable in terms of the random effects (see [Kenny et al., 2006](#)). This means that the random effects in the physiological linkage models were constrained to be the same across both members of the dyad (see the SM for more information on the random effects). A three-level condition variable compares ST targets, ST non-targets, and participants in control dyads. We use the term "respondent" to denote predictor variables for one person that affect the same person's outcome variable. We use the term "partner" to denote predictor variables from a partner. The structuring of the data (as a person period pairwise file, see [West, 2013](#)) allows us to estimate partner effects for both interaction partners in the model.

7.1.5.1. Questions asked and performance. Generalized Estimating Equations (GEE) were used to analyze questions asked about math (specifying a Poisson distribution with a log function because the dependent variable is count data) and performance (specifying a binomial distribution for wrong/correct with a logit function because the dependent variable is binary). GEE models allow us to adjust for non-independence over time and between dyad members ([Ballinger, 2004](#); [Liang & Zeger, 1986](#); [Zeger & Liang, 1986](#)). Because they can model outcomes that are exponentially distributed (e.g., count or binary), they are ideal for analyzing the number of questions asked (a count outcome) and performance (a binary outcome). Test statistics allow researchers to test hypotheses regarding parameter estimates in GEE that are analogous to those used in testing coefficients in regression and repeated measures ANOVA ([Rotnitzky & Jewell, 1990](#)).

7.1.5.2. Physiological linkage. To examine whether ST targets showed physiological linkage from one 30 s interval to the next to partners who talked more about math, we estimated a *stability and influence model* ([Thorson et al., 2018](#); [West, Shelton, & Trail, 2009](#)) with PROC MIXED in SAS ([West et al., 2017](#)). Participants' PEP reactivity at one point was treated as a function of their own reactivity at the prior time point (30 s prior, the stability path) and their partner's reactivity at that prior time point (the linkage path). Degrees of freedom are estimated using the Satterthwaite method, which involves a weighted average of the

between-subjects and within-subjects degrees of freedom (see Fitzmaurice, Laird, & Ware, 2011; Kenny et al., 2006). Degrees of freedom in this method, which can be fractional, are based on the total number of data points and adjusted for the nonindependence of observations. Because the nonindependence of observations is taken into account, the degrees of freedom in these analyses vary across different tests.

To examine whether ST targets attuned to partners who talked about math more, we ran a model in which the stability and linkage paths were moderated both by condition and talk time about math. Following the recommendations of Ledermann and colleagues (Ledermann, Macho, & Kenny, 2011), this model was fully saturated in that it contained both respondent and partner behavior as moderators of both stability and linkage; this strategy allowed us to isolate the effects of partner behavior on linkage while adjusting for any empirical overlap between respondent and partner behavior, and to look at linkage while adjusting for stability, which provides a conservative estimate of linkage. The interaction of theoretical interest is the three-way partner prior PEP by partner talk time by condition interaction. This interaction tests whether the strength of the linkage path varies by condition and as a function of how much time the partner talks about math.

7.2. Results

7.2.1. Social engagement with female peers

Recall that we predicted that women who were threatened prior to the interaction would show more social engagement—that is, asking more questions about the math task and showing stronger physiological linkage to her partner when that partner was talking about math—than women who were not threatened in these dyads (ST non-targets) and women in dyads in which neither partner was threatened (controls).

7.2.1.1. Questions asked about math task. A main effect of condition, Wald $\chi^2(2) = 36.31, p < .001$, indicated that ST targets ($M = 1.77, SD = 1.50$) asked significantly more questions about math (per 3-question segment) than ST non-targets did ($M = 1.14, SD = 1.29$), Wald $\chi^2(1) = 32.62, p < .001$, and controls ($M = 1.30, SD = 1.27$), Wald $\chi^2(1) = 18.16, p < .001$. Controls and ST non-targets did not significantly differ from one another, Wald $\chi^2(2) = 2.42, p = .36$.

7.2.1.2. Physiological linkage. Participants were stable from one moment to the next, as indicated by a main effect of respondent prior PEP on respondent current PEP, $b = 0.36, SE = 0.03, t(34.20) = 13.83,$

Table 2

All main effects and interactions from the stability and influence model used for physiological linkage in Study 1A.

| | Numerator df | Denominator df | F | p |
|---|--------------|----------------|--------|--------|
| Respondent prior PEP | 1 | 34.20 | 191.39 | < .001 |
| Partner prior PEP | 1 | 57.40 | 10.72 | .002 |
| Condition | 2 | 52.90 | 0.89 | .42 |
| Respondent prior PEP × Condition | 2 | 51.40 | 0.42 | .66 |
| Partner prior PEP × Condition | 2 | 59.80 | 1.15 | .33 |
| Respondent talk time | 1 | 2875 | 11.31 | .001 |
| Partner talk time | 1 | 2801 | 6.75 | .01 |
| Respondent talk time × Condition | 2 | 2691 | 0.82 | .44 |
| Partner talk time × Condition | 2 | 2715 | 0.05 | .95 |
| Respondent prior PEP × Respondent talk time | 1 | 2461 | 3.12 | .08 |
| Respondent prior PEP × Partner talk time | 1 | 2089 | 0.12 | .73 |
| Respondent prior PEP × Respondent talk time × Condition | 2 | 2592 | 0.17 | .84 |
| Respondent prior PEP × Partner talk time × Condition | 2 | 2105 | 2.59 | .07 |
| Partner prior PEP × Respondent talk time | 1 | 1517 | 0.16 | .69 |
| Partner prior PEP × Partner talk time | 1 | 1832 | 0.83 | .36 |
| Partner prior PEP × Respondent talk time × Condition ^a | 2 | 1371 | 5.43 | .005 |
| Partner prior PEP × Partner talk time × Condition | 2 | 1830 | 3.13 | .04 |

Note. The dependent variable is respondent current PEP. df = degrees of freedom.

^a We describe this effect in the SM.

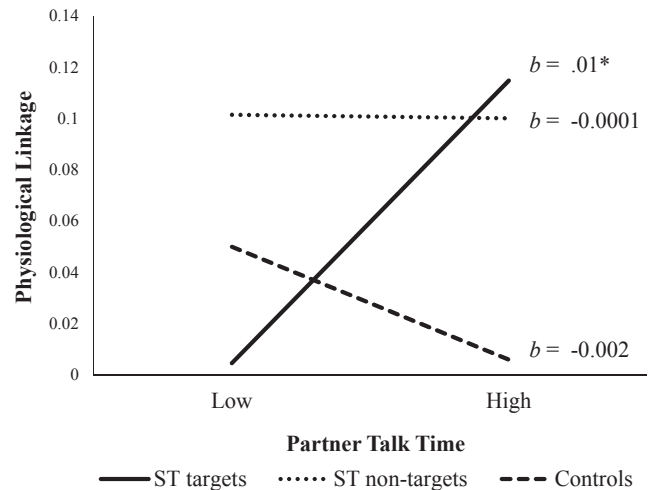


Fig. 5. Study 1A: Physiological linkage as a function of condition and partner talk time about math. Predicted values are presented at ± 1 standard deviation from the mean for partner talk time about math. * $p < .05$. The Y value is the unstandardized effect of the partner's prior PEP score on the respondent's current PEP score.

$p < .001$, 95% CI: 0.31–0.41 (see Table 2 for all main effects and interactions in the model). The main effect of partner prior PEP on respondent current PEP was also significant, indicating that participants experienced physiological linkage to their partners overall, $b = 0.06, SE = 0.02, t(57.40) = 3.27, p = .002$, 95% CI: 0.02 to 0.10. This effect was qualified by a Partner Prior PEP × Condition × Partner Talk Time interaction, $F(2, 1830) = 3.13, p = .044$, indicating that the strength of the linkage paths were different by condition, and further differed by how much one's partner talked about math. ST targets did not show linkage to their partners overall, $b = 0.06, SE = 0.03, t(57.60) = 1.84, p = .07$, 95% CI: -0.01 to 0.12. However, a significant two-way Partner Prior PEP × Partner Talk Time interaction was found for ST targets: ST targets were more physiologically linked to their partners (ST non-targets) during times when ST non-targets were talking more about math, $b = 0.01, SE = 0.003, t(1855) = 2.07, p = .039$, 95% CI: 0.0003–0.01 (see Fig. 5).

ST non-targets showed significant linkage to their partners (ST targets) overall, $b = 0.10, SE = 0.03, t(56.1) = 3.03, p = .004$, 95% CI: 0.03–0.17, but this effect was not moderated by partner talk time,

$b = -0.0001$, $SE = 0.002$, $t(1248) = -0.04$, $p = .97$, 95% CI: -0.004 to 0.004 . Controls did not show significant linkage to their partners (other control participants) overall, $b = 0.03$, $SE = 0.03$, $t(63.3) = 0.80$, $p = .43$, 95% CI: -0.04 to 0.10 , nor was the linkage path for controls moderated by partner talk time, $b = -0.002$, $SE = 0.002$, $t(2695) = -1.42$, $p = .16$, 95% CI: -0.01 to 0.001 . Thus, only ST targets showed physiological linkage to their partners (ST non-targets) when those partners talked more about math.

7.2.2. Performance

As a reminder, we proposed that when women work with women, women who had undergone a stereotype threat experience would not perform worse than those who had not undergone the stereotype threat manipulation, given our theorizing that these women would not be in a threat state when doing the math task. There was no main effect of condition, Wald $\chi^2(2) = 0.11$, $p = .95$, nor a significant main effect for time, $b = 0.01$, $SE = 0.002$, Wald $\chi^2(1) = 3.77$, $p = .052$, 95% CI: -0.00005 to 0.01 . In addition, we did not find a significant two-way Condition \times Time interaction, Wald $\chi^2(2) = 5.11$, $p = .078$. For interested readers, we report the individual effects of time for ST targets, ST non-targets, and controls in the SM. In brief, ST targets and controls improved over time, but ST non-targets did not. However, we caution interpreting these effects given that the overall interaction between Condition and Time was not significant.

7.2.3. Summary

Consistent with hypotheses from a *challenge and threat* perspective, ST targets who interacted with women displayed a behavioral pattern consistent with challenge by exhibiting the most social engagement—they asked the most questions about math and showed the strongest physiological linkage when their partners were talking about math. They also performed comparably to controls and ST non-targets. The fact that ST non-targets showed consistently high linkage to ST targets (i.e., linkage that was not conditional on the amount of time the ST targets spend talking about math) suggests that ST targets were consistently capturing the attention of their partners (the ST non-targets). This may have occurred, at least in part, because the ST targets were asking a lot of questions, which successfully drew their partners in to an interaction with them. Indeed, recent work has shown that when people capture the attention of others during group decision-making, other people show linkage to them (Thorson et al., 2018). Although these ST non-targets may have given a lot of attention to the ST targets, this attention did not seem to be directly focused on STEM content (i.e., it was not conditional on ST target talk time about math). For this reason, we do not interpret this finding as one showing that ST non-targets were either more or equally engaged in STEM-based discussions. Instead, they seemed generally attentive to their partners, but not in a way that was focused on STEM.

8. Study 1B

Study 1B ($n_{dyads} = 70$, $n_{participants} = 140$) employed the same design as Study 1A, except women always engaged with a male partner. This study was not combined into a single study with Study 1A because the only way to analyze the data from a combined study, while taking into account all of the dimensions on which participants differ (respondent gender, partner gender, respondent condition, and partner condition), is with a 7-level condition variable. This approach would make it difficult to see trends that occur within groups that share some characteristics in common (e.g., what happens among stereotype threat dyads vs. control dyads or what happens only among people who had female partners). However, following the presentation of results from Study 1B (in Section 9), we analyze the data from only the stereotype threat dyads across both studies to examine the experiences of women who underwent the stereotype threat manipulation when they were paired with females versus males. We also do the same using the data

from the control dyads to examine the experiences of women in the control condition when they were paired with females versus males.

8.1. Method

8.1.1. Design

Dyads either had one woman who received the ST manipulation (“ST targets”; $n = 40$) and one male participant who received the control manipulation (“ST non-targets”; $n = 40$; see Fig. 3), which we call “stereotype threat dyads”, or one woman and one man who both received the control manipulation (“female controls”; $n = 30$; “male controls”; $n = 30$), which we call “control dyads.”

8.1.2. Participants

Participants were math-identified college students who each had knowledge of the stereotype that men are better at math than women (140 participants; 56 Asian or Pacific Islander, 49 White, 19 Hispanic, 4 Black, 8 multiracial, 3 “other” and 1 “unknown”; $M_{age} = 19.97$, $SD_{age} = 1.43$); seventeen dyads who are not included in the dyads above were excluded *a priori* (thirteen for experimenter error and four for participant non-compliance).

8.1.3. Procedure and measures

The procedure and measures of Study 1B were the same as in Study 1A. Participants were pre-screened to make sure they were each 1) aware of the stereotype that men are better at math than women ($M = 1.88$, $SD = 0.66$) and 2) highly identified with math ($\alpha = 0.68$; $M = 5.85$, $SD = 0.55$). As in Study 1A, after training, one coder overlapped with each additional coder for 10% of the videos. Interrater reliability was assessed using a one-way random effects single-measures ICC (McGraw & Wong, 1996). The resulting ICCs were in the excellent range (ICC for questions asked = 0.87; ICC for talk time = 0.95; Cicchetti, 1994), indicating that questions asked and talk time were coded similarly across coders.

8.1.4. Analytic strategy

We used the same analytic strategy as in Study 1A. We examined the main effect of gender (male vs. female), condition of the dyads (ST dyad vs. control dyad), and a gender by condition interaction term in all models.

8.2. Results

8.2.1. Social engagement with male peers

When women who have undergone a stereotype threat experience work with men, recall that we proposed they would show similar or less social engagement compared to those who have not undergone the stereotype threat experience, depending on whether men serve to exacerbate the threat experience or lead to threat for all women.

8.2.1.1. Questions asked about math task. There was no main effect of gender, Wald $\chi^2(1) = 1.77$, $p = .18$, condition, Wald $\chi^2(1) = 2.01$, $p = .16$, nor a Condition \times Gender interaction, Wald $\chi^2(1) = 0.02$, $p = .89$; ST targets: $M = 1.50$, $SD = 1.56$; ST non-targets: $M = 1.38$, $SD = 1.52$; female controls: $M = 1.62$, $SD = 1.44$; male controls: $M = 1.51$, $SD = 1.43$.

8.2.1.2. Physiological linkage. Participants were stable from one moment to the next, as indicated by a main effect of respondent prior PEP on respondent current PEP, $b = 0.37$, $SE = 0.02$, $t(48.8) = 16.83$, $p < .001$, 95% CI: 0.33–0.42 (see Table 3 for all main effects and interactions in the model). The main effect of partner prior PEP on respondent current PEP was not significant, indicating that participants did not experience physiological linkage to their partners overall, $b = 0.02$, $SE = 0.02$, $t(87.7) = 1.21$, $p = .23$, 95% CI: -0.01 to 0.05 .

The two-way Partner Prior PEP \times Partner Talk Time interaction was

Table 3
All main effects and interactions from the stability and influence model used for physiological linkage in Study 1B.

| | Numerator df | Denominator df | F | p |
|--|--------------|----------------|--------|--------|
| 1. Respondent prior PEP | 1 | 48.8 | 283.16 | < .001 |
| 2. Partner prior PEP | 1 | 87.7 | 1.48 | .23 |
| 3. Gender | 1 | 50.4 | 0.94 | .34 |
| 4. Condition | 1 | 50.5 | 0.80 | .38 |
| 5. Gender × Condition | 1 | 50.4 | 4.99 | .03 |
| 6. Respondent prior PEP × Gender | 1 | 48.8 | 14.05 | < .001 |
| 7. Respondent prior PEP × Condition | 1 | 48.8 | 3.78 | .06 |
| 8. Respondent prior PEP × Gender × Condition | 1 | 48.8 | 0.38 | .54 |
| 9. Partner prior PEP × Gender | 1 | 86.2 | 5.42 | .02 |
| 10. Partner prior PEP × Condition | 1 | 87.7 | 0.07 | .79 |
| 11. Partner prior PEP × Gender × Condition | 1 | 86.2 | 1.06 | .31 |
| 12. Respondent talk time | 1 | 3629 | 0.21 | .65 |
| 13. Partner talk time | 1 | 3552 | 1.73 | .19 |
| 14. Respondent talk time × Gender | 1 | 3609 | 0.02 | .88 |
| 15. Respondent talk time × Condition | 1 | 3629 | 0.35 | .56 |
| 16. Respondent talk time × Gender × Condition | 1 | 3609 | 0.02 | .89 |
| 17. Partner talk time × Gender | 1 | 3542 | 0.72 | .40 |
| 18. Partner talk time × Condition | 1 | 3552 | 1.08 | .30 |
| 19. Partner talk time × Gender × Condition | 1 | 3542 | < 0.01 | .99 |
| 20. Respondent prior PEP × Respondent talk time | 1 | 1753 | 1.73 | .19 |
| 21. Respondent prior PEP × Partner talk time | 1 | 1494 | 0.39 | .53 |
| 22. Respondent prior PEP × Respondent talk time × Gender | 1 | 1599 | < 0.01 | .95 |
| 23. Respondent prior PEP × Respondent talk time × Condition | 1 | 1753 | 1.64 | .20 |
| 24. Respondent prior PEP × Respondent talk time × Gender × Condition | 1 | 1599 | 5.97 | .01 |
| 25. Respondent prior PEP × Partner talk time × Gender | 1 | 1327 | 0.27 | .60 |
| 26. Respondent prior PEP × Partner talk time × Condition | 1 | 1494 | 0.87 | .35 |
| 27. Respondent prior PEP × Partner talk time × Gender × Condition | 1 | 1327 | 0.20 | .65 |
| 28. Partner prior PEP × Respondent talk time | 1 | 564 | < 0.01 | .96 |
| 29. Partner prior PEP × Partner talk time | 1 | 788 | 1.63 | .20 |
| 30. Partner prior PEP × Respondent talk time × Gender | 1 | 569 | 1.19 | .28 |
| 31. Partner prior PEP × Respondent talk time × Condition | 1 | 564 | 0.82 | .37 |
| 32. Partner prior PEP × Respondent talk time × Gender × Condition | 1 | 569 | 0.02 | .90 |
| 33. Partner prior PEP × Partner talk time × Gender | 1 | 794 | 0.19 | .66 |
| 34. Partner prior PEP × Partner talk time × Condition | 1 | 788 | 0.07 | .79 |
| 35. Partner prior PEP × Partner talk time × Gender × Condition | 1 | 794 | < 0.01 | .99 |

Note. The dependent variable is respondent current PEP. *df* = degrees of freedom.

not significant, $F(1, 788) = 1.63, p = .20$, indicating that the strength of the linkage path did not differ by partner talk time about math. This interaction was not moderated by condition, gender, or an interaction of the two ($ps > 0.66$). Unlike in Study 1A, in which ST targets linked more strongly to female partners when they were talking about math, we did not find that they linked more strongly to male partners when they were talking about math.

8.2.2. Performance

As a reminder, we proposed that if male partners exacerbate the threat experience for women who have undergone stereotype threat, then women under stereotype threat would perform worse than (a) women who were not put under stereotype threat and (b) men. However, if all women experience threat when working with men, regardless of their stereotype threat condition, we proposed that all women would perform worse than men performed.

We found a significant main effect of gender, Wald $\chi^2(1) = 41.39, p < .001$, such that females performed worse than males performed overall (see Fig. 6). In addition, there was a two-way Gender × Time interaction, Wald $\chi^2(1) = 7.85, p = .005$, such that females improved over time, $b = 0.01, SE = 0.003, Wald \chi^2(1) = 5.18, p = .02, 95\% CI: 0.001-0.011$, but males did not, $b = -0.001, SE = 0.003, Wald \chi^2(1) = 0.30, p = .58, 95\% CI: -0.01 to 0.004$. At the beginning of the task, males performed significantly better than females did, Wald $\chi^2(1) = 27.7, p < .001$, but there was no effect of gender at the end of the task, Wald $\chi^2(1) = 0.49, p = .48$. There was no main effect of condition, no Condition × Time interaction, and no Condition × Gender × Time interaction ($ps > 0.15$).

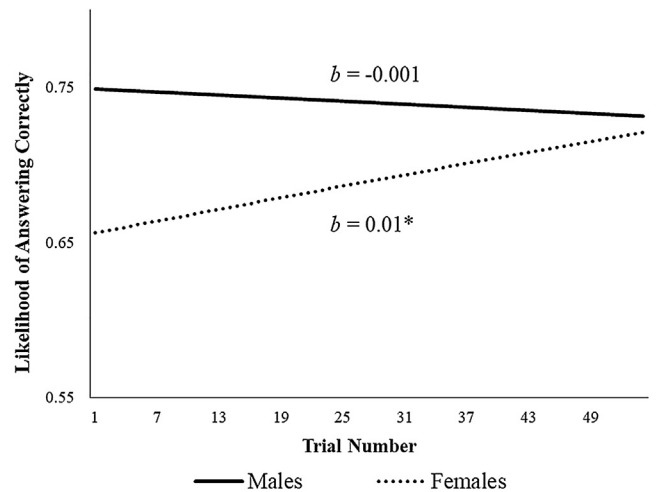


Fig. 6. Study 1B: Likelihood of answering a question correctly as a function of gender across time. $^*p < .05$.

8.2.3. Summary

When working with male partners, stereotype threatened women did not engage more with their male partners than control women did: they did not ask more questions and they were not linked to their partners on the basis of how much they talked about math. We also found that women—regardless of stereotype threat condition—performed worse than men performed.

9. Cross-study comparisons

Because Studies 1A and 1B are identical methodologically except for the gender of the partner, we combined the results across both studies to facilitate conclusions about the effect of partner gender on women’s social engagement in dyadic STEM contexts. Specifically, we examined how threatened women’s social engagement and performance varied as a function of their partner’s gender and whether or not similar patterns for social engagement and performance as a function of partner gender emerged among women who were not threatened. We note that these comparisons are not experimental because random assignment to condition did not occur for these comparisons.

9.1. Social engagement

9.1.1. Stereotype threat dyads

First, we analyzed results from the stereotype threat dyads in both studies, resulting in four types of participants: ST targets who interacted with female ST non-targets (from Study 1A), female ST non-targets who interacted with ST targets (from Study 1A), ST targets who interacted with male ST non-targets (from Study 1B), and male ST non-targets who interacted with ST targets (from Study 1B).

We examined the impact of a 4-level variable representing the four types of participants listed above on the number of questions participants asked their partners about math. We observed a significant effect of condition, $Wald \chi^2(3) = 32.79, p < .001$ (see Fig. 7). This effect was driven by the difference between female ST targets paired with women, who asked more questions than their partners did (female non-ST targets; $p < .001$, as reported in Study 1A) and more questions than male non-ST targets did ($p = .006$). In addition, female non-ST targets paired with female ST targets asked fewer questions than female ST targets paired with men did ($p = .006$). The primary comparison of interest—comparing ST targets with female partners to ST targets with male partners—was not significant, $Wald \chi^2(1) = 5.10, p = .14$, meaning that women under threat did not ask more questions of female partners than male partners did. None of the other pairwise comparisons were significant ($ps > 0.14$).

We also examined the impact of a 4-level variable representing the four types of participants listed above on physiological linkage as a function of partner talk time about math. We did not observe a significant effect of condition on this outcome, $F(3, 2063) = 1.06, p = .36$.

9.1.2. Control dyads

Next, we analyzed results from the control dyads in both studies, resulting in three types of participants: female controls who interacted with female controls (from Study 1A), female controls who interacted with male controls (from Study 1B), and male controls who interacted with female controls (from Study 1B).

We examined the impact of a 3-level variable representing the three types of participants listed above on the number of questions participants asked their partners about math. We observed a significant effect

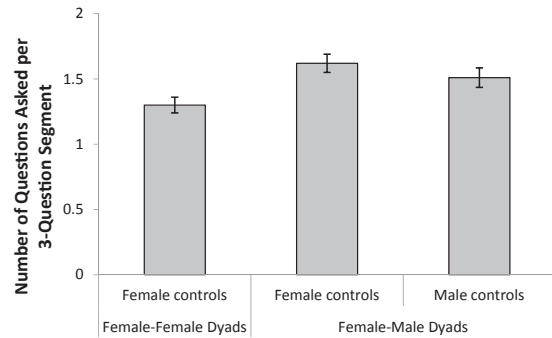


Fig. 8. Number of questions asked per 3-question segment among control dyads only. Error bars represent 95% confidence intervals.

of condition, $Wald \chi^2(2) = 8.39, p = .015$ (see Fig. 8). The pairwise comparison of interest—comparing female controls who interacted with female controls to female controls who interacted with male controls—was significant, $Wald \chi^2(1) = 8.06, p = .014$, indicating that control women asked more questions of male partners than of female partners. Note that in the hypotheses section (under female-male dyads), we predicted that if men served as a stressor, then women should be less engaged with men than with women; this effect is in the opposite direction of that predicted effect. None of the other pairwise comparisons were significant ($ps > 0.17$).

We also examined the impact of a 3-level variable representing the four types of participants listed above on physiological linkage as a function of partner talk time about math. There was no effect of condition on this outcome, $F(2, 471) = 1.16, p = .32$, indicating that none of the control groups were linked to their partners based on how much they talked.

9.2. Performance

9.2.1. Stereotype threat dyads

Similar to the analyses for social engagement, we examined the impact of a 4-level variable representing the four types of participants listed above on performance. We found a significant effect of condition, $Wald \chi^2(3) = 40.01, p < .001$ (see Fig. 9). This effect was driven by male non-ST targets who performed better than their female partners, better than female ST targets paired with female non-ST targets, and better than female non-ST targets paired with female ST targets ($ps < 0.001$). The primary comparison of interest, however—comparing ST targets with female partners to ST targets with male partners—was not significant, $p = .77$, meaning that women under stereotype threat did not differ in performance with female vs. male partners. None of the other pairwise comparisons were significant ($ps > .18$).

We also found a significant interaction of Condition \times Time, $Wald \chi^2(3) = 10.66, p = .014$. This effect was driven by the difference in performance over time between female ST targets paired with males

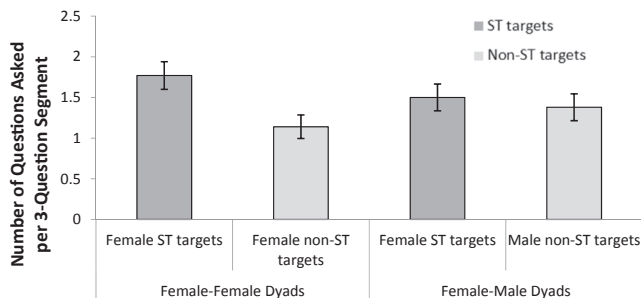


Fig. 7. Number of questions asked per 3-question segment among stereotype threat dyads only. Error bars represent 95% confidence intervals.

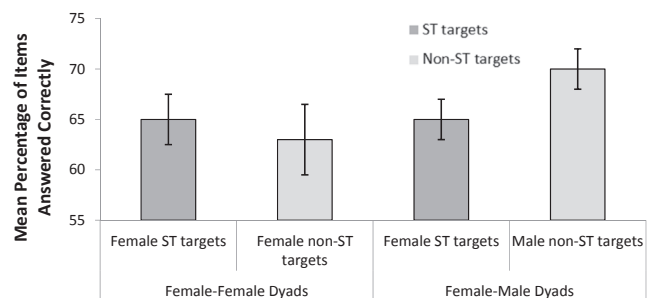


Fig. 9. Mean percentage of items answered correctly among stereotype threat dyads only. Error bars represent 95% confidence intervals.

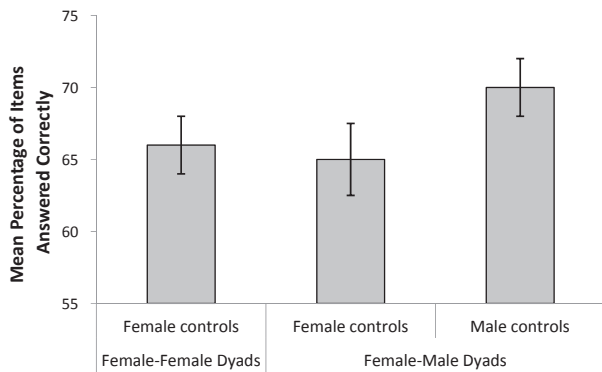


Fig. 10. Mean percentage of items answered correctly among control dyads only. Error bars represent 95% confidence intervals.

(who improved in performance over time) relative to female non-ST targets paired with females (who did not change in performance over time, $p = .031$) and male non-ST targets paired with females (who did not change in performance over time, $p = .008$). The pairwise comparison of interest, however—comparing ST targets with female partners to ST targets with male partners—was not significant ($p = .43$), meaning that the performance over time of women under stereotype threat did not differ with female vs. male partners. None of the other pairwise comparisons were significant ($ps > .07$).

9.2.2. Control dyads

Similar to the analyses for social engagement, we examined the impact of a 3-level variable representing the three types of participants listed above on performance. We found a significant effect of condition, Wald $\chi^2(2) = 21.13$, $p < .001$ (see Fig. 10). This effect was driven by male controls paired with female partners, who performed better than female controls paired with female controls ($p = .003$) and female controls paired with male controls ($p = .001$). The pairwise comparison of interest, however—comparing female controls who interacted with female controls to female controls who interacted with male controls—was not significant, Wald $\chi^2(1) = 1.06$, $p = .91$.

9.3. Summary

Taken together with the findings of Studies 1A and 1B, the cross-study comparisons reveal that overall, men did not serve as a stressor that led to or perpetuated the stereotype threat experience for women. When comparing across women who underwent the threat manipulation, those who were subsequently paired with men were no less likely to engage with their partners than those who were paired with women. In addition, among women who did not undergo threat, those who interacted with men actually showed some evidence of stronger engagement (question asking) than those who interacted with women. Overall, then, the biggest difference in how experiencing stereotype threat prior to a dyadic interaction affected women is within same-gender (female-female) interactions, where women who experienced the threat manipulation prior to the interaction engaged more with their partners than those who did not.

10. Discussion

There are a number of key findings in the present research. First, consistent with our hypotheses, we found that experiencing stereotype threat prior to working with a female partner led to more social engagement for women, relative to those who did not experience stereotype threat prior to an interaction with a female partner. Specifically, women under stereotype threat asked female partners more questions and were more physiologically linked to them when they were discussing math than women in dyads where neither partner was under

threat. Second, we found that in female-female dyads, stereotype threat did not affect performance: those who experienced stereotype threat performed comparably to those who did not. Third, we did not find evidence that men served as stressors for women. We found that experiencing stereotype threat prior to working with a male partner lead neither to more nor less social engagement for women (question asking, physiological linkage) relative to those who did not experience stereotype threat prior to an interaction with a male partner and relative to those who experienced threat prior to an interaction with a female partner. Fourth, in female-male dyads, we found that women performed worse than men did regardless of whether they had experienced stereotype threat prior to the interaction.

Taken together, the biggest effects we observed regarding how the experience of stereotype threat affects behavior were within female-female interactions, comparing women who experienced threat to those who did not. These findings have important theoretical implications for how best to address the issue of stereotype threat in collaborative learning environments. To date, much of the emphasis on the role of gender in STEM interactions has been on how to structure groups—including the ratio of women to men and the frequency of interaction between women to men—with the goal of preventing or dampening the experience of stereotype threat for women. Some scholars have suggested that the presence of men might fuel the psychological experience of stereotype threat for women, and one way to combat this experience is to encourage more frequent interactions between women in STEM (e.g., increase collaborative learning among women and increase the number of female role models for women; e.g., Dasgupta et al., 2015; Marx & Roman, 2002). Applications of this approach are common within STEM-based scholarship programs, which not only encourage girls and women to form relationships with other female peers but also to develop relationships with female mentors (e.g., The Girls' Network in the United Kingdom). Our research suggests that in contexts where female-female interactions are taking place, it is critical to understand what experiences women have prior to these interactions. To the extent that women have experienced stereotype threat prior to their interactions with women, they might engage with other women in ways that are quite different than women who have not.

Perhaps one of the most surprising findings is that we found very little evidence that male interaction partners either operated as stereotype threat “inducers” or exacerbated the experience of threat for women who had already undergone a threat manipulation. We did find that women performed worse than men performed overall (in Study 1B), but we are hesitant to interpret these findings as evidence of stereotype threat, given that women who were paired with a male interaction partner did not perform worse than women paired with a female interaction partner did (overall across conditions). Indeed, we found that women actually performed slightly better when with men (mean percentage of items answered correctly = 65.4) than when with women (mean percentage of items answered correctly = 64.9), Wald $\chi^2(1) = 3.49$, $p = .062$. In other words, performance differences between men and women cannot be attributed to the gender of one's interaction partner. Why women performed worse than men overall remains an open question for future research. It may be the case that performance differences in a highly evaluative context are due to chronic experiences of stereotype threat—a hypothesis that future research could test.

Why didn't male partners serve as stressors for women? One possibility is that male partners were not stressors because they were not in an evaluative role when they interacted with women—instead they were equal-status peers who were also participants in the research study. It is also possible that male partners were not threatening because they were a stark contrast to the male evaluators. Future research should examine engagement for women who undergo a stereotype-threatening experience and then interact with a higher-status male (e.g., a professor or a tutor), as well as how back-to-back interactions with men of different roles affect women's experiences.

Our findings highlight the importance of considering social context in understanding when and how the presence of men will lead women to experience stereotype threat. They also have implications for the debate about whether coeducational or single-sex classrooms are better environments to promote women's participation in STEM (see Billger, 2009; Cherney & Campbell, 2011; Pahlke & Hyde, 2016). Our results align with a recent meta-analysis of Grades K-12, showing no difference in a host of outcomes, including mathematics performance, attitudes towards mathematics, science performance, and educational aspirations in coeducational versus single-sex classrooms (Pahlke, Hyde, & Allison, 2014), suggesting that male peers do not universally induce stereotype threat.

10.1. Educational implications

We have argued that because much of STEM learning and work occurs in dyads and small groups, the practical implications of our work extend to those in STEM learning and working contexts. In STEM contexts where women work with women—such as female-only study groups or in single-sex classrooms—prior experiences of stereotype threat may boost social engagement for women. In this sense, our work reveals an unforeseen beneficial effect of stereotype threat (a phenomenon for which researchers have shown almost exclusively negative to neutral effects; for a review across domains, see Schmader et al., 2008): improving social engagement for women around women. Specifically, we proposed that the heightened arousal brought about by math stereotype threat (Ben-Zeev et al., 2005; Mendes & Jamieson, 2012; O'Brien & Crandall, 2003; Schmader et al., 2008) can lead to a behavioral pattern of challenge when women are around other women that manifests as greater engagement with others. We did not find that experiencing stereotype threat improved performance for women under threat who worked with female partners, but this may be because peers vary in their level of expertise and so attending to peers and asking them questions is unlikely to lead to better performance for everyone who engages in these processes. Future research might vary the level of expertise of the partner and test how engagement with partners who provide valuable feedback fosters performance (Grover et al., 2017). For example, in contexts where learning particular strategies for solving problems is useful (Star, Caronongan, Foegen, Furgeson, Keating, Larson, & Zbiek, 2015), engagement with expert peers who can teach those strategies might be more likely to boost performance.

Although we did not find an effect of stereotype threat on women's performance with other women, the results we found regarding social engagement are still relevant to important outcomes, such as keeping women in STEM. In fact, recent work has shown that better grades among undergraduate female engineering students are not associated with retention in engineering majors or with career aspirations in engineering (Dennehy & Dasgupta, 2017). However, feelings of social belongingness (such as feeling connected to peers) are associated with retention and career aspirations. In sum, this work suggests that better performance does not keep students in STEM but feeling accepted and engaged with other students does. Thus, even though the results found here do not demonstrate changes in women's performance, they do show changes in another critical process for keeping women in STEM: being socially engaged with others.

There are several implications for being physiologically linked to other people in educational settings. Research has shown that people are more likely to be physiologically linked to interaction partners who are good at capturing their attention, either by engaging in motivationally relevant behaviors (such as displaying cues of anxiety) or by being persuasive (Kraus & Mendes, 2014; Thorson et al., 2018; West et al., 2017). In educational contexts in particular, greater physiological and neurological linkage is associated with shared attention (e.g., when a teacher is lecturing; Dikker et al., 2017). Thus, to the extent that students are paying attention to someone who is helping them or providing useful information (which we aimed to capture by examining

physiological linkage while people's partners were speaking about math), being physiologically linked to others may have performance benefits in the long run (a possibility which we discuss in the Future Directions section below). Second, being physiologically linked to others when they are discussing STEM-relevant content reflects engagement with others who are engaged with STEM material in particular (and not social engagement more generally). Because we know from prior research that engagement with other people in STEM fields, like engineering, can buffer undergraduates from failures in STEM and promote STEM career aspirations (Dennehy & Dasgupta, 2017; Solder et al., 2012; Walton et al., 2015), this type of physiological linkage may also provide similar benefits in the long run.

Outside of female-only contexts and in the majority of STEM education and work settings where women work with men, our research suggests that experiencing stereotype threat prior to working with a man has no additional impact on women's engagement or performance when compared with not experiencing stereotype threat prior to the interaction. Therefore, efforts to eliminate or mitigate the psychological experience of stereotype threat are important in all mixed-gender STEM contexts and not just ones that feature frequent explicit social evaluation.

The performance patterns that we observed for women who interacted with men may have implications for how STEM classroom activities are structured. We found that women improved in performance over the course of a 30-minute interaction, starting off worse than men but performing comparably by the end of the task. If these effects are due to women becoming more comfortable with their male partners over time, they suggest that women may be best-served by consistent work with the same male partners (e.g., having the same lab partners throughout the semester instead of switching every week). We note that our work was limited to a 30-minute interaction only, and thus, future work should examine whether these patterns replicate in longer-term settings (like with lab partners over the course of the semester) and in other educational settings.

10.2. Limitations and future directions

There are important limitations of the present research and avenues for future research. One question that remains is the extent to which the types of engagement we measured are associated with selecting a STEM education or career. Dyadic social engagement may help women establish social networks within STEM that buffer them from adversity (Walton et al., 2015) and ultimately prevent them from dropping out of STEM. In this research, we measured both an overt form of engagement (asking questions) and a more automatic, undetectable one (conditional physiological linkage). Future research might consider which of these types of engagement is tied to the selection of a STEM education or career path, as well as how engaging with males versus females affects whether women stay in STEM. Consistent with research examining undergraduates in engineering, engaging with other women might be particularly protective against STEM dropout (Dasgupta et al., 2015). Furthermore, research could consider whether the difference in the perceptibility of these behaviors plays a role. For example, a student who asks her teacher a lot of questions is likely to receive more encouragement from her teacher than a student who is silently, but closely, attending to the teacher's behaviors. Understanding how dyadic engagement begets commitment to STEM careers, and the process through which this association occurs, is a critical step towards preventing STEM drop-out.

It is also important to understand how engagement affects performance. We did not predict that asking questions or attending to the partner when they were talking about math would necessarily lead to better performance because partners varied in the quality of their answers and the content of their math discussions. Engaging with competent peers or higher-status people, like professors or bosses, should be beneficial for performance because women can gain useful information

via engagement and also signal an interest in the field to others, who may then be more likely to provide support and mentorship. It is also possible that initially engaging with a new peer can improve performance, as engaging with the new peer allows one to figure out whether the peer is a useful source of information. If they are not useful, students may learn to rely on themselves more than their peers, which could also potentially improve their performance. Future research should examine how engagement with competent and incompetent partners affects performance and the time course of these effects.

Given the role that physiological stress reactivity has been shown to play in math stereotype threat effects (Ben-Zeev et al., 2005; Mendes & Jamieson, 2012; O'Brien & Crandall, 2003; Schmader et al., 2008) future work might address whether physiological threat reactivity to stereotype threat also plays a role in the dyadic outcomes assessed in this study. For example, it is possible that women who experience the most physiological threat in response to experiencing stereotype threat are those who are the most likely to engage with a female partner. Assessing physiological threat while women experience the stereotype threat manipulation and then linking that to dyadic outcomes could help reveal how stereotype threat affects engagement with a new partner.

We note that researchers have studied the effects of stereotype threat by using a range of methods to induce stereotype threat experiences. The method we used here combined the use of a diagnostic math task and the presence of males to prime concerns about STEM aptitude and one's gender within a social evaluation paradigm widely used in the stress and coping literature—the Trier Social Stress Test (Kirschbaum et al., 1993). In addition, the gender of experimenters was varied across conditions—with male experimenters in the stereotype threat condition and female experimenters in the control condition. Scholars have theorized that making the stereotype salient is essential to evoking the experience of threat, and we attempted to do this by having male evaluators. However, because the present manipulation involves multiple components, it is not clear which of these components is essential for evoking the threat responses we observed for women. For instance, it might be the case that women experience stereotype threat when they are evaluated by higher-status women, as these women may evoke concerns about confirming stereotypes (e.g., concerns about “letting other women down”; J.R. Shapiro, 2011). Future work might tease apart these different components to figure out which elements, or combination of elements, have the greatest impact on students' social engagement within STEM. Such research would have important implications for how women engage with others in several social situations, including interactions within female mentoring programs. In such programs, women entering STEM fields are often directly evaluated by those who are higher-status (e.g., female professors for college students, female managers in a pharmaceutical firm).

In this research, we were specifically interested in how stereotype threat affects the social engagement of female students who were highly identified with math, as these are the students most likely to pursue STEM careers (although we did not directly measure whether they were math or STEM majors or intending to pursue careers in math or STEM) and are, therefore, good targets for interventions aimed at preventing STEM disengagement. In addition, we reasoned that women who are not highly identified with math are unlikely to be affected by negative stereotypes about women's math performance (Aronson et al., 1999; Keller, 2007; Steele, 1997). The patterns observed in this research might not necessarily extend to students who are less identified with math, and future research might examine when and how engagement with STEM might be improved for these students and potentially get them “back on track” to pursue careers in STEM. All of the participants in this research were students at the same school (New York University), and it is possible that the patterns in this research might not extend to students at different types of schools—for example, schools with large engineering programs where students might commonly experience unbalanced gender ratios.

Finally, Asians are the majority racial group in all of our samples, and research has shown that stereotypes about math can affect female Asian students in contrasting ways. When the gender of female Asians is salient, stereotype threat effects emerge, presumably because women are associated with worse performance in math; however, stereotype threat effects do not occur when the race of female Asians is salient, presumably because Asians are associated with better—and not worse—performance in math (Shih, Pittinsky, & Ambady, 1999). Therefore, by activating the gender-math stereotype and not the race-math stereotype, our manipulation of stereotype threat likely affected both Asian and non-Asian women similarly. Because our samples were primarily composed of White and Asian students, future research should examine how female students who are under-represented minorities would also be affected.

10.3. Conclusion

Gender gaps within STEM remain a problem. In the present research, we investigated how stereotype threat affects women's engagement when working with partners during a dyadic math task. We found that experiencing stereotype threat prior to working with a female partner led to more social engagement for women (but no differences in performance), relative to those who did not experience stereotype threat prior to an interaction with a female partner. When working with male partners, we found that women performed worse than men did regardless of whether they had experienced stereotype threat prior to the interaction. Our work reveals novel dyadic engagement processes that are a result of experiencing stereotype threat and provides implications for a variety of working contexts within STEM.

Acknowledgements

The authors would like to thank the research assistants from The West Interpersonal Perception Lab at NYU for their dedication to this research. Some of the data in this manuscript were presented at the Society for Experimental Social Psychology (2015) and the Society for Personality and Social Psychology (2016). This work was supported by a Grant-in-Aid from the Society for the Psychological Study of Social Issues (awarded to Katherine R. Thorson) and by the National Science Foundation (grant number DRL1535414, awarded to Chad E. Forbes and Tessa V. West).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cedpsych.2019.03.012>.

References

- Abdi, H. (2007). The Bonferroni and Šidák corrections for multiple comparisons. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 103–107). Thousand Oaks, CA: Sage.
- Accreditation Board for Engineering and Technology (2017). Accreditation criteria and supporting docs. Retrieved on March 7, 2017 from <http://www.abet.org/accreditation/accreditation-criteria/>.
- Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45(5), 369–386. <https://doi.org/10.1002/pits.20303>.
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35(1), 29–46. <https://doi.org/10.1006/jesp.1998.1371>.
- Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods*, 7, 127–150. <https://doi.org/10.1177/1094428104263672>.
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, 41(2), 174–181. <https://doi.org/10.1016/j.jesp.2003.11.007>.
- Berntson, G. G., Lozano, D. L., Chen, Y. J., & Cacioppo, J. T. (2004). Where to Q in PEP. *Psychophysiology*, 41(2), 333–337.
- Billger, S. M. (2009). On reconstructing school segregation: The efficacy and equity of

- single-sex schooling. *Economics of Education Review*, 28(3), 393–402. <https://doi.org/10.1016/j.econedurev.2007.08.005>.
- Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology*, 35(1), 61–79. [https://doi.org/10.1016/S0022-4405\(96\)00029-5](https://doi.org/10.1016/S0022-4405(96)00029-5).
- Blascovich, J., & Mendes, W. B. (2010). Social psychophysiology and embodiment. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 194–227). (5th ed.). New York, NY: John Wiley & Sons Inc.
- Blascovich, J., Vanman, E., Mendes, W. B., & Dickerson, S. (2011). *Social psychophysiology for social and personality psychology*. Sage Publications.
- Bolger, N., Stadler, G., & Laurenceau, J. P. (2011). Power analysis for intensive longitudinal studies. In M. R. Mehl, & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 285–301). New York, NY: Guilford.
- Capraro, R. M., Capraro, M. M., & Morgan, J. R. (2013). *STEM project-based learning: An integrated science, technology, engineering, and mathematics (STEM) approach*. Boston, MA: Sense Publishers.
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135(2), 218–261. <https://doi.org/10.1037/a0014412>.
- Cherney, I. D., & Campbell, K. L. (2011). A league of their own: Do single-sex schools increase girls' participation in the physical sciences? *Sex Roles*, 65(9–10), 712. <https://doi.org/10.1007/s11199-011-0013-6>.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>.
- Crocker, J., Luhtanen, R. K., Cooper, M. L., & Bouvrette, A. (2003). Contingencies of self-worth in college students: Theory and measurement. *Journal of Personality and Social Psychology*, 85(5), 894–908. <https://doi.org/10.1037/0022-3514.85.5.894>.
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardt, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28(12), 1615–1628. <https://doi.org/10.1177/014616702237644>.
- Dasgupta, N., Scirle, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences*, 112(16), 4988–4993. <https://doi.org/10.1073/pnas.1422822112>.
- Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences*, 114(23), 5964–5969. <https://doi.org/10.1073/pnas.1613117114>.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3), 355–391. <https://doi.org/10.1037/0033-2909.130.3.355>.
- Dikker, S., Wan, L., Davidesco, I., Kaggan, L., Oostrik, M., McClintock, J., ... Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, 27(9), 1375–1380. <https://doi.org/10.1016/j.cub.2017.04.002>.
- Dovidio, J. F. (2001). On the nature of contemporary prejudice: The third wave. *Journal of Social Issues*, 57(4), 829–849. <https://doi.org/10.1111/0022-4537.00244>.
- Drane, D., Smith, H. D., Light, G., Pinto, L., & Swarat, S. (2005). The gateway science workshop program: Enhancing student performance and retention in the sciences through peer-facilitated discussion. *Journal of Science Education and Technology*, 14(3), 337–352. <https://doi.org/10.1007/s10956-005-7199-8>.
- Finnigan, K. M., & Corker, K. S. (2016). Do different types of stereotype threat moderate the effect of performance-avoidance goals on female math performance? *Journal of Research in Personality*, 63, 36–43. <https://doi.org/10.1016/j.jrp.2016.05.009>.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons Inc.
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of Social Issues*, 53(1), 25–44.
- Forbes, C. E., Amey, R., Magerman, A. B., Duran, K., & Liu, M. (2018). Stereotype-based stressors facilitate emotional memory neural network connectivity and encoding of negative information to degrade math self-perceptions among women. *Social Cognitive and Affective Neuroscience*, 13(7), 719–740.
- Forbes, C. E., Duran, K. A., Leitner, J. B., & Magerman, A. (2015). Stereotype threatening contexts enhance encoding of negative feedback to engender underperformance and anxiety. *Social Cognition*, 33(6), 605–625. <https://doi.org/10.1521/soco.2015.33.6.605>.
- Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of Personality and Social Psychology*, 99(5), 740–754. <https://doi.org/10.1037/a0020971>.
- Fox, S. I. (2006). *Human physiology* (3rd ed.). Boston, MA: McGraw Hill.
- Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, 1(4), 379–390. <https://doi.org/10.1037/1082-989X.1.4.379>.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafo, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>.
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology*, 49(10), 1886. <https://doi.org/10.1037/a0031412>.
- Gasiewski, J. A., Eagan, M. K., Garcia, G. A., Hurtado, S., & Chang, M. J. (2012). From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Research in Higher Education*, 53(2), 229–261. <https://doi.org/10.1007/s11162-011-9247-y>.
- Gibson, C. E., Losee, J., & Vitiello, C. (2014). A replication attempt of stereotype susceptibility (Shih, Pittinsky, & Ambady, 1999): Identity salience and shifts in quantitative performance. *Social Psychology*, 45(3), 194. <https://doi.org/10.1027/1864-9335/a000184>.
- Grover, S. S., Ito, T. A., & Park, B. (2017). The effects of gender composition on women's experience in math work groups. *Journal of Personality and Social Psychology*, 112(6), 877–900. <https://doi.org/10.1037/pspi0000090>.
- Hall, W., Schmader, T., & Croft, E. (2015). Engineering exchanges: Women's daily experience of social identity threat in engineering cue burnout. *Social Psychological and Personality Science*, 6(5), 528–534. <https://doi.org/10.1177/1948550615572637>.
- Heyman, R. E., Lorber, M. F., Eddy, J. M., & West, T. V. (2013). Behavioral observation and coding. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 451–477). New York, NY: Cambridge University Press.
- Holleran, S. E., Whitehead, J., Schmader, T., & Mehl, M. R. (2011). Talking shop and shooting the breeze: A study of workplace conversation and job disengagement among STEM faculty. *Social Psychological and Personality Science*, 2(65), 65–71. <https://doi.org/10.1177/1948550610379921>.
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, 11(5), 365–371. <https://doi.org/10.1111/1467-9280.00272>.
- Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology*, 46(1), 208–212. <https://doi.org/10.1016/j.jesp.2009.08.015>.
- Jamieson, J. P., Nock, M. K., & Mendes, W. B. (2012). Mind over matter: Reappraising arousal improves cardiovascular and cognitive responses to stress. *Journal of Experimental Psychology: General*, 141(3), 417–422.
- Jamieson, J. P., Peters, B. P., Greenwood, E. J., & Altose, A. J. (2016). Reappraising stress arousal improves performance and reduces evaluation anxiety in classroom exam situations. *Social Psychological and Personality Science*, 7(6), 579–587. <https://doi.org/10.1177/1948550616644656>.
- Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology General*, 137(4), 691–705. <https://doi.org/10.1037/a0013834>.
- Jones, M. G., Howe, A., & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education*, 84(2), 180–192. [https://doi.org/10.1002/\(SICI\)1098-237X\(200003\)84:2<180::AID-SCE3>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1098-237X(200003)84:2<180::AID-SCE3>3.0.CO;2-X).
- Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, 77(2), 323–338. <https://doi.org/10.1348/000709906X113662>.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.
- Kirschbaum, C., Pirke, K., & Hellhammer, D. H. (1993). The "Trier Social Stress Test": A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1–2), 76–81. <https://doi.org/10.1159/000119004>.
- Knapp, T. R. (2015). In (partial) defense of .05. *Journal of Modern Applied Statistical Methods*, 14(2), 27–34. <https://doi.org/10.22237/jmasm/1446350700>.
- Kraus, M. W., & Mendes, W. B. (2014). Sartorial symbols of social class elicit class-consistent behavioral and physiological responses: A dyadic approach. *Journal of Experimental Psychology*, 143(6), 2330–2340. <https://doi.org/10.1037/xge0000023>.
- Lane, S. P., & Hennes, E. P. (2018). Power struggles: Estimating sample size for multilevel relationships research. *Journal of Social and Personal Relationships*, 35(1), 7–31.
- Lazarus, R., & Folkman, S. (1991). The concept of coping. In A. Monat, & R. S. Lazarus (Eds.), *Stress and coping: An anthology* (pp. 189–206). (3rd ed.). New York: Columbia University Press.
- Ledermann, T., Macho, S., & Kenny, D. A. (2011). Assessing mediation in dyadic data using the actor-partner interdependence model. *Structural Equation Modeling*, 18(1), 595–612. <https://doi.org/10.1080/10705511.2011.607099>.
- Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, 63(2), 234–246.
- Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22. <https://doi.org/10.1093/biomet/73.1.13>.
- Liu, S., Rovine, M. J., Cousino Klein, L., & Almeida, D. M. (2013). Synchrony of diurnal cortisol pattern in couples. *Journal of Family Psychology*, 27(4), 579. <https://doi.org/10.1037/a0033735>.
- Logel, C., Peach, J. M., & Spencer, S. J. (2012). Threatening gender and race: different manifestations of stereotype threat. In M. Inzlicht, & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 159–172). New York, NY: Oxford University Press. <https://doi.org/10.3389/fpsyg.2015.00900>.
- Lozano, D. L., Norman, G., Knox, D., Wood, B. L., Miller, B. D., Emery, C. F., & Berntson, G. G. (2007). Where to B in dZ/dT. *Psychophysiology*, 44(2007), 113–119. <https://doi.org/10.1111/j.1469-8986.2006.00468.x>.
- Major, B., & Schmader, T. (1998). Coping with stigma through psychological disengagement. In J. K. Swim, & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 219–241). <https://doi.org/10.1016/B978-012679130-3/50045-4>.
- Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, 28(9), 1183–1193. <https://doi.org/10.1177/01461672022812004>.
- McEwen, B. S. (2007). Physiology and neurobiology of stress and adaptation: Central role of the brain. *Physiological Reviews*, 87(3), 873–904. <https://doi.org/10.1152/physrev.00041.2006>.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass

- correlation coefficients. *Psychological Methods*, 1(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>.
- Mendes, W. B. (2016). Emotion and the autonomic nervous system. In L. E. Barrett, M. Lewis, & J. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 166–181). (4th ed.). New York, NY: Guilford Press.
- Mendes, W. B., Gray, H. M., Mendoza-Denton, R., Major, B., & Epel, E. S. (2007). Why egalitarianism might be good for your health: Physiological thriving during stressful intergroup encounters. *Psychological Science*, 18(11), 991–998. <https://doi.org/10.1111/j.1467-9280.2007.02014.x>.
- Mendes, W. B., & Jamieson, J. (2012). Embodied stereotype threat: Exploring brain and body mechanisms underlying performance impairments. In M. Inzlicht, & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 51–68). New York, NY: Oxford University Press.
- Micari, M., Pazos, P., Streitwieser, B., & Light, G. (2010). Small-group learning in undergraduate STEM disciplines: Effect of group type on student achievement. *Educational Research and Evaluation*, 16(3), 269–286. <https://doi.org/10.1080/13803611.2010.520860>.
- Murphy, M., Steele, C., & Gross, J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. Retrieved from *Psychological Science*, 18(10), 879–885. <http://www.jstor.org/stable/40064666>.
- National Science Foundation (2017). Women, minorities, and persons with disabilities in science and engineering: 2017. Retrieved on March 7, 2017 from <https://nsf.gov/statistics/2017/nsf17310/>.
- Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334. <https://doi.org/10.1037/a0012702>.
- O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, 29(6), 782–789. <https://doi.org/10.1177/0146167203252810>.
- Pahlke, E., & Hyde, J. S. (2016). The debate over single-sex schooling. *Child Development Perspectives*, 10(2), 81–86. <https://doi.org/10.1111/cdep.12167>.
- Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, 140(4), 1042. <https://doi.org/10.1037/a0035740>.
- Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2016). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review*, 1–43. <https://doi.org/10.1177/1088868316628405>.
- Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. R. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLOS ONE*, 11(1), e0146487. <https://doi.org/10.1371/journal.pone.0146487>.
- Rotnitzky, A., & Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77, 485–497. <https://doi.org/10.1093/biomet/77.3.485>.
- Saxbe, D., & Repetti, R. L. (2010). For better or worse? Coregulation of couples' cortisol levels and mood states. *Journal of Personality and Social Psychology*, 98(1), 92. <https://doi.org/10.1037/a0016959>.
- Schachinger, H., Weinbacher, M., Kiss, A., Ritz, R., & Langewitz, W. (2001). Cardiovascular indices of peripheral and central sympathetic activation. *Psychosomatic Medicine*, 63(5), 788–796. <https://doi.org/10.1097/00006842-200109000-00012>.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38(2), 194–201. <https://doi.org/10.1006/jesp.2001.1500>.
- Schmader, T., Forbes, C. E., Zhang, S., & Mendes, W. B. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin*, 35(5), 584–596. <https://doi.org/10.1177/0146167208330450>.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440. <https://doi.org/10.1037/0022-3514.85.3.440>.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336–356. <https://doi.org/10.1037/0033-295X.115.2.336>.
- Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology*, 39(1), 68–74. [https://doi.org/10.1016/S0022-1031\(02\)00508-5](https://doi.org/10.1016/S0022-1031(02)00508-5).
- Selye, H. (1946). The general adaptation syndrome and the diseases of adaptation. *The Journal of Clinical Endocrinology*, 6(2), 117–230.
- Shapiro, E. S. (2011). *Academic skills problems* (4th ed. workbook). New York, NY: The Guilford Press.
- Shapiro, J. R. (2011). Different groups, different threats: A multi-threat approach to the experience of stereotype threats. *Personality and Social Psychology Bulletin*, 37(4), 464–480. <https://doi.org/10.1177/0146167211398140>.
- Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology*, 104(2), 277. <https://doi.org/10.1037/a0030461>.
- Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., & Van Doornen, L. J. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology*, 27(1), 1–23.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80–83. <https://doi.org/10.1111/1467-9280.00111>.
- Solder, M., Rowan-Kenyon, H., Inkelas, H. K. H., Garvey, J., & Robbins, C. (2012). Supporting students' intentions to persist in STEM disciplines: The role of living learning programs among other social-cognitive factors. *The Journal of Higher Education*, 83(3), 311–336. <https://doi.org/10.1353/jhe.2012.0017>.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28. <https://doi.org/10.1006/jesp.1998.1373>.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21–51. <https://doi.org/10.2307/1170643>.
- Star, J. R., Caronongan, P., Foegen, A. N., Furgeson, J., Keating, B., Larson, M. R., ... Zbiek, R. M. (2015). Teaching strategies for improving algebra knowledge in middle and high school students. Education Books. Retrieved from the NCEE website: <http://whatworks.ed.gov>.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34, 379–440. [https://doi.org/10.1016/S0065-2601\(02\)80009-0](https://doi.org/10.1016/S0065-2601(02)80009-0).
- Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, 100(2), 255–270. <https://doi.org/10.1037/a0021385>.
- Tien, L. T., Roth, V., & Kampmeier, J. A. (2002). Implementation of a peer-led team learning instructional approach in an undergraduate organic chemistry course. *Journal of Research in Science Teaching*, 39(7), 606–632. <https://doi.org/10.1002/tea.10038>.
- Thorson, K. R., Dumitru, O. D., Mendes, W. B., & West, T. V. (2019). Status enhances approach behaviors during group decision-making, but leadership and competence are needed to successfully persuade others (in preparation).
- Thorson, K. R., West, T. V., & Mendes, W. B. (2018). Measuring physiological influence in dyads: A guide to designing, implementing, and analyzing dyadic physiological studies. *Psychological Methods*, 23(4), 595–616.
- Trawalter, S., Richeson, J. A., & Shelton, J. N. (2009). Predicting behavior during interracial interactions: A stress and coping approach. *Personality and Social Psychology Review*, 13(4), 243–268.
- U.S. Department of Commerce (2013). Disparities in STEM employment by sex, race, and Hispanic origin. Retrieved on March 7, 2017 from <https://census.gov/prod/2013pubs/acs-24.pdf>.
- Vick, S. B., Seery, M. D., Blascovich, J., & Weisbuch, M. (2008). The effect of gender stereotype activation on challenge and threat motivational states. *Journal of Experimental Social Psychology*, 44(3), 624–630. <https://doi.org/10.1016/j.jesp.2007.02.007>.
- Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a “chilly climate” transform women's experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, 107(2), 468–485. <https://doi.org/10.1037/a0037461>.
- Waters, S. F., West, T. V., & Mendes, W. B. (2002). Stress contagion: Physiological covariation between mothers and infants. *Psychological Science*, 25(4), 934–942. <https://doi.org/10.1177/0956797613518352>.
- Waters, S. F., West, T. V., Karnilowicz, H. R., & Mendes, W. B. (2017). Affect contagion between mothers and infants: Examining valence and touch. *Journal of Experimental Psychology: General*, 146(7), 1043–1051. <https://doi.org/10.1037/xge0000322>.
- West, T. V. (2013). Repeated measures with dyads. In J. A. Simpson, & L. Campbell (Eds.), *The Oxford handbook of close relationships* (pp. 731–749). New York: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195398694.013.0033>.
- West, T. V., Koslov, K., Page-Gould, E., Major, B., & Mendes, W. B. (2017). Contagious anxiety: Anxious European Americans can transmit their physiologic reactivity to African Americans. *Psychological Science*, 28(12), 1796–1806. <https://doi.org/10.1177/0956797617722551>.
- West, T. V., Shelton, J. N., & Trail, T. E. (2009). Relational anxiety in interracial interactions. *Psychological Science*, 20(3), 289–292. <https://doi.org/10.1111/j.1467-9280.2009.02289.x>.
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039. <https://doi.org/10.1126/science.1136099>.
- Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121–130. <https://doi.org/10.2307/2531248>.